THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 38 March, 1947 Number 3

THE PREDICTION OF INTELLIGENCE AT COLLEGE ENTRANCE FROM EARLIER TEST*

ROBERT L. THORNDIKE
Teachers College, Columbia University

PURPOSE OF THE STUDY

Intelligence testing has been going on in the schools of our country for about a generation now. Some schools have been carrying on an extensive program of testing for a good many years, and of course the output of research studies based upon this testing has been considerable. But it seemed to the author that there were still opportunities to bring the accumulated test data in the files of schools and of testing organizations to bear upon problems significant for psychology and education. The problems which are attacked in this investigation are:

(1) How accurately can intelligence at the time of college entrance (defined as performance upon a specific test) be predicted from tests given at various earlier ages?

(2) With what weights should two or more tests given at different times during a child's school career be combined to give the most accurate prediction of intelligence at the time of college entrance?

* Funds for this study were provided by a grant from the Carnegie Corporation of New York.

The study was made possible by the coöperation of many persons and agencies. The College Entrance Examination Board made a major contribution in permitting use of data on its Scholastic Aptitude Test as the terminal measure of ability. Collecting data on earlier tests was greatly facilitated by the coöperation of the Educational Records Bureau, which made available the central file of testing which it had carried out for its subscribing schools. A number of schools gave generously of their time in supplying lists of pupils and data concerning those pupils. These schools are listed at the end of the article.

SUMMARY AND INTERPRETATION OF RESULTS

The procedure for this investigation was to determine the extent of correlation between a terminal test given at the end of the secondary school (the College Entrance Examination Board Scholastic Aptitude Test) and intelligence tests given at various intervals prior to the terminal test. The standard error of estimate of terminal from earlier test was also computed in each case. All available test data were assembled for about five thousand pupils who had taken the terminal test and for whom earlier test data were available. Over thirteen thousand test records were analysed in addition to the terminal tests. These tended to be concentrated in the secondary-school years, and data for tests in the elementary grades were relatively meager. However, data were obtained for over six hundred tests administered ten or more years prior to the terminal test.

The data indicate that the drop in accuracy of prediction as one goes from a test administered in the same year as the terminal test to one administered two, four, six, or even eight years earlier is small, irregular, and possibly non-existent. It can be stated with fair assurance that time interval is not a significant factor in determining the accuracy of prediction for secondary-school tests. Although the data are more meager for the upper elementary grades, they suggest that the same situation holds there too. In the case of the primary grades, it appears that correlations with the terminal test are substantially lower, though this may be in large measure a reflection of difference in the nature of the tests.

These findings have several significant implications for educational practice. They imply that insofar as the factor of intelligence is concerned, we can give a pupil substantially as accurate guidance concerning his future educational plans at the beginning of his high-school course, and perhaps even in the upper elementary grades, as we can near the end of the course. It implies that if a high school plans to give an intelligence test for use in educational guidance in and by the high school, that test should be given early in the high-school period. The results can be expected to be as accurate for prognosis as those from a later test, and will be available for use for a longer period of time. It implies that if more than one test has been administered to a

pupil during his high-school career, and probably including the upper elementary-school years, the best prediction of his college-age intelligence is obtained from a simple average of the two test scores. (This assumes that the two tests are equally good. The present study provides no basis for evaluating the worth of different tests.) Averaging two or more scores will yield a more reliable measurement and, if different tests are used, will tend to broaden somewhat the concept of intelligence upon which the resulting score is based. It implies that tests, and especially group tests, in the primary grades should not be used for long-range forecasting, but should be supplemented by later tests.

ASSEMBLING THE DATA

The general plan of this study was to locate a group of individuals for whom schools had records of intelligence tests at various times through their school career and for whom a terminal test was available at or near the level of intellectual maturity. The latest date at which it seemed practical to find tests on a large population was college entrance, so this was chosen as the time for the terminal test. A uniform test given under specified conditions and uniformly standardized seemed very desirable for this final test, and so approach was made to the College Entrance Examination Board for permission to use the accumulated data upon its Scholastic Aptitude Test. This permission was generously granted.

The Scholastic Aptitude Test had a number of advantages for our study. One very important practical advantage was that the files of the Board contained, at the time the study was initiated, test results for some one hundred fifty thousand pupils arranged in compact and accessible form. This made it possible to locate terminal records for our original group of about ten thousand individuals with a minimum expenditure of time and effort. The efficient clerical service supplied by the Board reduced what originally loomed as a formidable obstacle to a few weeks' work.

The more fundamental advantage of the Scholastic Aptitude Test is that it is a carefully constructed, homogeneous, relatively 'pure' test carefully standardized so that the scores have uniform meaning from year to year. The results are reported in standard scores (mean = 500, standard deviation = 100) based on the

population of ten to fifteen thousand who take the test annually. The comparability of different editions has been checked by including overlapping material in the examinations and by pretesting the material of new issues. A possible limitation of the S.A.T. lies in its very homogeneity. It measures a somewhat limited range of verbal ability (there is also a separate mathematical section in some cases but this was not available for enough cases to be useable in this investigation), and probably cannot be considered synonymous with 'intelligence' in any broad and inclusive sense. Perhaps we should uniformly speak of it as verbal ability, but it will be more convenient to drop the qualifying 'verbal' in our discussion and speak simply of 'ability' or 'aptitude' or 'intelligence.' Whenever we are speaking of terminal intelligence in this report is should be understood to mean 'verbal score on the College Entrance Examination Board Scholastic Aptitude Test.'

Once the Scholastic Aptitude Test had been chosen as the terminal test, the next task was to locate pupils who had taken this test and for whom there existed a fairly rich record of earlier testing. From the reports of the Board, schools were located which (1) had contributed a fair number of pupils to the Board testing and (2) were located within a reasonable distance of New York City. To these a letter and questionnaire were sent explaining the study, soliciting coöperation, and asking for data concerning the extent of testing within the school. Over fifty schools finally met the criteria of being willing to coöperate and of having sufficient earlier test records to make their inclusion in the study seem worth while. A number of other schools expressed a willingness to coöperate, but were not finally included in the study for one reason or another.

Each school was asked to supply a list of the graduates from the school who had taken the S.A.T. from 1936 to 1940, inclusive. In some cases, data were also assembled for individuals who had taken the test in 1935 or in 1941. Each name from these lists was transcribed on to a data card, and the file of data cards was sent to the C.E.E.B. for the entry of S.A.T. score. At this point, the file contained about ten thousand names.

After the S.A.T. score had been recorded on the data cards, the cards were returned, and a research worker went to each coöperating school to transcribe earlier records for each individual

on to the data card. The record of each test was transcribed in as complete detail as feasible, so as to make possible subsequent checking of ages and computed mental ages and IQ's. This stage of the work was much expedited by making use of the facilities of the Educational Records Bureau. Many of the coöperating schools were members of the Bureau. With the permission of the schools and of the Bureau, it was possible to assemble much of the data on earlier testing from the Bureau's files. Other records were gathered directly from the files of the coöperating schools.

A major problem that was encountered in the study was the large amount of shifting of pupils from one school to another. This was exaggerated by the fact that a number of the schools which were on other counts the most desirable were college-preparatory of the type to which many boys or girls go for only the last year or two before college. It became necessary, therefore, to try to trace pupils back to earlier schools. The terminal schools (and the E.R.B.) were drawn upon once more for data on previous schools attended by the pupils. For a number of pupils it was impossible to get this information, and for these we could go no further. In most cases, however, information on earlier schools was obtained, and it was possible to refer back to those schools for additional data on the individuals.

Records from earlier schools were, of course, widely scattered. From some schools there were quite substantial groups, while from many schools there were only one or two cases. Explanatory letters and record blanks were sent out to five hundred thirty-eight of the most promising schools. The response was most gratifying. Replies were received from three hundred ninety-nine schools. Not all of these contributed additional data, but on the whole a substantial addition was made to the test records by including data from the earlier schools. Time did not permit making this follow-up of earlier records complete, but the most readily available of the records were obtained.

At this point, we still had in our files record cards for some ten thousand individuals. Some of the cards contained nothing but a name. Some contained records of as many as ten or twelve intelligence tests, scattered over a dozen years of the pupil's school career and including a Scholastic Aptitude Test given in the junior or senior year of high school as a part of the

procedure of college entrance. A first step was to reduce the number of cases by eliminating those for whom the data were most inadequate. Certain test records were eliminated because they were incomplete, because they appeared to be in error, or because the test in question had norms of a type which were not adapted to this study. Then the population of cases was reduced by eliminating (1) all cases for whom there was no terminal S.A.T. and (2) all cases for whom there was not, in addition, either one test more than four years prior to the S.A.T. or at least two tests in different years prior to the S.A.T. This reduced the population to some five thousand cases. These cases provided the material for further analysis.

The next step was to check over all the test records, catching obvious errors and inconsistencies, and reducing all occurrences of the same test to comparable form. For most of the tests, the results were reported in terms of IQ's computed from the most recent norms for the test. Old test records which would have used earlier norms were converted into the IQ's which would have been received according to the current norms. Where the data were too incomplete to effect such a conversion, or the appropriateness of the conversion was in doubt, the record was discarded. A large body of testing for which IQ's were not used was E.R.B. testing using the American Council on Education Psychological Examination. These records, which were generally reported as raw scores and/or as percentile ranks for the particular grade in the E.R.B. Independent Schools population, were translated into T-scores for the Independent Schools population, assuming a normal distribution with a mean of 50 and a standard deviation of 10. The population tested by the E.R.B. is believed to be quite stable in character from year to year, and from grade to grade in the grades using the A.C.E. examination, so these grade standard scores may be expected to have a satisfactorily uniform meaning in terms of the degree of excellence represented by a particular score from grade to grade and from year to year.

After all the records had been checked and reduced either to IQ's or standard scores, they were coded and punched on Hollerith cards. In dealing with the data, it seemed essential to keep separate the records for different intelligence tests. Since the variable in which we were primarily interested was interval between prior test and terminal test, this was the basis for segre-

gating different fields in the punch card. Only one test record could be entered in a given field (set of columns in the card), so it was necessary to reject extra tests when more than one test had been given to an individual within a particular time interval. Data were classified in terms of years before the terminal test, and only one test could be recorded in a given year. An arbitrary hierarchy of preference was established to determine which test should be kept. In general, the aim was to get a substantial body of data for certain frequently occurring tests, at the expense of other more fragmentary data. Subsequent analyses of the data were made from the coded punched card records.

ANALYSIS OF RESULTS

There were four different times in the school careers of secondary-school students at which they might take the College Entrance Examination Board Scholastic Aptitude Test. The test was given both in the junior and in the senior years, and both in April and in June. These different times of testing were required to serve the needs of different schools and colleges and also of competitive scholarship awards. The variations in time of testing made it necessary to break down the available records before further analyses were carried out. Four subgroups were set up based upon the time of this terminal test as follows:

Group I	Tested in April of junior year
Group II	Tested in June of junior year
Group III	Tested in April of senior year
Group IV	Tested in June of senior year

It turned out that the number of cases in Group I was too small for profitable analysis, so these cases were discarded and analyses were carried through separately for each of the three remaining groups.

Within each of the three subgroups which were analyzed, correlations were computed between the Scholastic Aptitude Test and all earlier tests for which sufficient data were available. The procedure was to compute separate correlations for each form of earlier test and for each time interval between earlier test and Scholastic Aptitude Test. That is, a correlation would be computed between Scholastic Aptitude Test and Otis Self Administrating Higher Examination, Form A with a one-year

TABLE I.—PREDICTION OF SCHOLASTIC APTITUDE TEST VERBAL SCORE RECORDS TABULATED BY PREDICTOR TEST AND BY TIME INTERVAL

	American Council Verbal Score				Otis S-A Higher				Otis S-A Intermediate			
Interval	N	r	S_{SAT}	S_{EST}	N	r	S_{SAT}	S_{EST}	N	r	S_{SAT}	S_{EST}
0	2609	.70	96.2	68.7	285	.64	98.2	75.5				
1	2258	.70	96.0	68.6	374	.65	96.4	73.3				
2	1373	.73	95.5	65.3	332	.62	90.9	71.3				
3	620	.69	95.2	68.9	329	.65	95.3	72.4	42	.48	99.7	87.4
4	90	.58	85.1	69.3	294	.65	82.4	62.6	123	.55	88.6	74.0
5					81	.73	107.6	73.6	226	.62	91.7	72.0
6									73	.58	72.4	59.0

	Otis Group Test				Terman Group Test				Henmon-Nelson High School			
Interval	N	r	S_{SAT}	S_{EST}	N	r	S_{SAT}	S_{EST}	N	r	S_{SAT}	S_{EST}
0	30	.73	94.2	64.4	26	.82	100.3	60.1				
1	47	.66	81.3	61.1	69	.81	94.2	55.2				
2	98	.72	92.5	64.2	202	.77	93.0	59.3	101	.70	107.6	76.9
3	83	.67	102.9	76.4	194	.69	95.5	69.1	154	.58	85.2	69.4
4					289	.77	97.8	62.4	146	.60	101.0	80.8
5					187	.71	90.7	63.9				
6					76	.71	87.7	61.8				

	Henmon-Nelson Elementary				Kuhlmann-Anderson				Haggerty Delta 2			
Interval	N	r	S_{SAT}	S_{EST}	N	r	S_{SAT}	S_{EST}	N	r	S_{SAT}	S_{EST}
2												
3					60	.61	89.2	70.7				
4					71	.48	90.8	79.6				
5	44	.69	96.9	70.1	144	.57	90.3	74.2	29	.69	78.6	56.9
6	100	.74	89.7	60.3	119	.32	88.3	83.7	93	.69	79.1	57.2
7												

TABLE I.—(*Continued*)

	National Intelligence Test				Stanford-Binet (1916)			
Interval	N	r	S_{SAT}	S_{EST}	N	r	S_{SAT}	S_{EST}
2					32	.71	91.0	64.1
3					52	.57	81.9	67.3
4					81	.44	98.5	88.4
5	31	.43	84.7	76.5	80	.56	101.5	84.1
6	104	.61	91.8	72.8	96	.55	91.3	76.3
7	127	.54	85.9	72.3	95	.59	94.3	76.1
8	167	.54	93.8	78.9	120	.67	92.6	68.7
9					64	.30	88.8	84.7
10					127	.49	104.0	90.7
11					173	.39	87.2	80.3
12					107	.39	87.9	81.0

time interval. Another correlation would be computed for that same test with a two-year interval. Another correlation would be computed for Form B of the Otis Test with the one-year interval, and so forth for all combinations of tests and intervals. However, correlations were only computed in those instances in which at least twenty-five individuals were available who had taken the particular test within the particular interval. It should be noted in passing that the one-year interval was defined as twelve to twenty-three months; the two-year interval as twenty-four to thirty-five months; and so forth.

In addition to all the correlations between Scholastic Aptitude Test and specific test forms, a certain number of correlations were also computed with grouping of forms of earlier tests. Thus, where there were not enough cases for computation with each form of the Otis Self-Administering Higher Examination, records from the several forms were thrown together and treated as the same test.

In this report it has seemed necessary to consolidate the complete table of correlations, because of its extreme bulk and unwieldiness. The consolidation was carried out by combining correlations for the three separate groups (Groups II III, and IV mentioned above) and for the several forms of essentially the same test. The combination was accomplished by trans-

lating each value of r into the corresponding value of Fisher's z, weighting by N-3 (the reciprocal of the variance of z), and averaging. The average z values were converted back to values of r. Consolidated values are also reported for the standard deviation of Scholastic Aptitude Test scores. These represent weighted averages of the standard deviations in the separate subgroups. The composite standard deviations were combined with the composite correlation values to yield a single value for the standard error of estimate for each test and interval. The complete table of correlations, together with N's, means and standard deviations, has been mimeographed and can be supplied by the author to anyone who is interested in it.*

The correlations of earlier tests with the terminal Scholastic Aptitude Test and the standard errors of estimate for predicting the Scholastic Aptitude Test are shown in Table I. A single entry in this table may represent the average of several values from different subgroups and different test forms. Thus, the entry for the Otis Self-Administering Higher Examination given one year before the terminal test is an average of 8 values for different subgroups and different forms of the Otis test.

Table I presents the essential results of this study with regard to the prediction of the terminal Scholastic Aptitude Test. It can be seen from examining the tables that the available data are ragged, leaving something to be desired in the way of completeness. Even though analyses were based on some five thousand individuals for whom records of the terminal test and one or more earlier tests were available, by the time these data had been broken down both by form of test and by interval between testings the resulting groups became quite small in many cases. The

---

* Two interesting technical problems arise in connection with the evaluation of these correlations. One concerns testing the comparability of the various subsamples upon which the several correlations are based. The other has to do with the differential effect of age upon scores of early and late tests. Some consideration was given to these problems. A relatively crude examination of the SAT variance in different subsamples suggested that variation among samples could probably be attributed to sampling fluctuations, and that no correction need be made for it. The reporting of standard errors of estimate partially takes care of the problem of differing variability in the subsamples. No way of allowing for differential age effects appeared wholly satisfactory, but it was not judged that these effects would appreciably distort the differences in correlations for different intervals.

available data are heavily concentrated in the two or three years immediately prior to the terminal test, and earlier records are relatively scanty. This is a natural situation when one is following back case records. It was accentuated in this instance by the fact that many of the schools studied were subscribers to the Educational Records Bureau and had carried out extensive and systematic testing during the secondary-school period. As a result quite extensive data are available particularly for the American Council on Education Psychological Examination administered in grades IX, X, XI, and XII. Though the earlier records were somewhat disappointing, it may still be valuable to bring together the records of several hundred Stanford-Binet Intelligence tests given at the kindergarten and first-grade level and to relate those to tests at the completion of the high-school period. In any event, in the interpretation of the data of Table I the most secure conclusions will relate to the years immediately preceding the terminal test and conclusions with regard to earlier tests will necessarily remain somewhat tentative.

Inspection of Table I suggests rather clearly that prediction of the terminal test is made about equally successfully by a test at any time during the high-school period. Particular attention should be paid to the results on the American Council Examination because of the large number of available cases. The results of this test suggest that prediction of the terminal Scholastic Aptitude Test can be made as successfully from a test administered in the ninth grade as from one administered only a few months before the terminal test. This conclusion is entirely supported by the results for the Otis Self-Administering Higher Examination. The only other test administered to substantial populations during the high-school period was the Terman Group Test of Intelligence. The results for this test are somewhat less clear-cut, and there is some slight indication of decline in prediction as the interval before the terminal test is increased. However, the groups in the zero and one-year interval are very small and the results for longer intervals are irregular. No other tests were given to enough individuals or over a broad enough span of intervals to contribute much to an understanding of prediction during this secondary-school period.

As we consider tests administered during the upper elementary grades, that is during the period four to eight years prior to the

terminal test, the available data become much less satisfactory. Correlations are available for a number of different tests, but in most cases these are based upon relatively small groups and cover a span of only a few years. It is possible to make a crude comparison of the correlations from these upper elementary-school tests with those obtained for the high-school tests. If an unweighted average is taken of all the standard errors of estimate for the one- and two-year intervals, for example, and compared

TABLE II.—CHANGE IN PREDICTABILITY OF TERMINAL TEST WITH INCREASE IN INTERVAL BETWEEN TESTS BASED ON ALL TESTS COMMON TO ADJACENT INTERVALS

Interval	Correlation Coefficient				Standard Error of Estimate			
	Average difference from previous interval		Cumulative difference from zero interval		Average difference from previous interval		Cumulative difference from zero interval	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
1 yr	−.018	.000	−.018	.000	−2.6	−0.4	−2.6	−2.6
2 yrs	+.005	+.017	−.013	+.017	+0.5	−2.6	−2.1	−3.0
3 yrs	−.067	−.038	−.080	−.021	+3.7	+3.6	+1.6	+0.6
4 yrs	−.026	−.001	−.106	−.022	+1.7	−1.6	+3.3	−1.0
5 yrs	+.060	+.037	−.046	+.015	+0.2	+0.4	+3.5	−0.6
6 yrs	−.010	−.057	−.056	−.042	−3.8	−2.5	−0.3	−3.1
7 yrs	−.015	−.017	−.071	−.059	−0.4	−0.4	−0.7	−3.5
8 yrs	+.040	+.038	−.031	−.021	−0.4	+0.6	−1.1	−2.9

with a similar unweighted average for the five- and six-year intervals, the values obtained are 65.9 and 69.5, respectively. These suggest a slight drop in accuracy of prediction for the longer interval. In this case, however, the effect of time interval is confused with the effect of varying test form. It appears that the nature of the particular test is a more potent determiner of correlations with the terminal test than is time interval. This can be seen by comparing different tests given during the same time period.

In an effort to clarify the effect of time interval, Table II was prepared. It shows the change in accuracy of prediction for each

year's increase in time interval. In addition, the cumulative change in accuracy of prediction is shown, obtained by summing these changes for adjacent years. Results are presented both for correlation coefficients and for standard errors of measurement, but attention should be directed chiefly to the latter, since they make allowance for differences in the variability of the subsamples. The values in the second and third columns of Table II give the average difference in correlation between two adjacent time intervals. They are based upon all the tests which were given to groups at both time intervals. Thus, the first entry, comparing tests given with zero and one-year intervals is based on the American Council, Otis SA Higher, Otis Group, and Terman Group tests. The last entry, comparing tests given with seven- and eight-year intervals, is based only upon the National and the Stanford-Binet tests. Column 2 gives unweighted averages; column 3 weights the differences for each test in proportion to the number of cases in the smaller of the two groups upon which the difference is based. (This is a rough approximation to weighting each difference in proportion to the reciprocal of its variance.) The weighted values should probably be given greater attention, because of the gross discrepancies in the size of the various samples. Columns 4 and 5 of the table show cumulative differences from an interval of zero years up to the interval indicated in column 1. That is, column 4 and 5 are cumulative values, summing the values in columns 2 and 3, respectively. Columns 6, 7, 8 and 9 show the analogous information for the standard errors of estimate. It should be noted that a negative change (decrease) in standard error of estimate represents an increase in accuracy of prediction.

From Table II we must conclude that there is no clear relationship between time interval and accuracy of prediction of terminal test score. The data for the correlation coefficients suggest a slight decline in accuracy of prediction as the time interval is increased, though the relationship is neither marked nor regular. However, in the case of the standard errors of estimate the relationship is, if anything, reversed. That is, there is a slight tendency for the standard error of estimate to become smaller, and the prediction correspondingly more accurate, as the interval is increased. Bearing in mind the limitations placed on any interpretation by the small groups in the elementary grades and the

shift in tests represented, the most reasonable interpretation of these data would seem to be that there is no relationship between time interval and accuracy of prediction of the terminal test over the range from 0 to 8 years.

Of particular interest are the sketchy results from testing eight or more years prior to the terminal test. This testing carries one back to the primary grades in elementary school and even to the kindergarten level. The possibility of predicting test performance at the end of high school from one of these tests given almost at the time the child enters school has a good deal both of practical and theoretical interest. Results for use of the Stanford Binet are shown in Table I. Binet records were available over a wide span of time intervals, but the groups were rather small and the correlations are correspondingly unstable. This fact appears clearly in an examination of the last column of Table I. Trends do not appear at all clearly in such unstable values as these. However, there appears to be some tendency for the prediction to become less accurate as the interval becomes greater. Paying attention to the tests administered ten or more years prior to the terminal test, we see that a value somewhere about .40 would probably represent the correlational data fairly well. This value is clearly a decided drop from the values obtained from group tests during the upper elementary period and from the values obtained with high-school tests.

In addition to the early Stanford Binet data, limited data were available for group tests administered ten to twelve years before the terminal test. Combining the several time intervals and the several subgroups, a correlation of .21 is obtained for one hundred thirty-one cases with the Pintner-Cunningham Primary Test and a correlation of .22 for eighty-three cases with the Detroit First Grade Intelligence Test. The standard deviations of the Scholastic Aptitude Test in these two cases are 96.0 and 96.7, respectively. These results suggest that the early group tests are definitely less satisfactory than the early Stanford Binet as far as predicting intelligence test performance at college entrance is concerned. The correlations from these early group tests appear almost vanishingly small, and they would certainly be of little value as a basis for long time predictions.

Some comment should perhaps be made on the general order of magnitude of the correlations in Table I. The highest cor-

relations with terminal test are those for the Terman Group Test, which approach .80. The American Council on Education Psychological Examination gives correlations about .70, and several other tests approach this level. Others tend to give correlations in the 60's or even lower. That these correlations are not higher is due in part to the imperfect reliability of the separate measures. No completely satisfactory reliability data are available for use with these correlations, but from retest data with a year interval it can be estimated that the reliability of the Scholastic Aptitude Test for a college applicant population is not far from .90. The reliabilities of the other tests are almost all certainly less. Thus, the correlation between ACE administrations a year apart was found in our data to be about .82. Using this and the value of .90 for the SAT, a correlation of .70 between the two tests becomes .81 when corrected for unreliability. Other correlations would be similarly affected, though the exact amount of the correction may be overestimated somewhat because of the nature of the reliability estimate.

Another factor reducing the correlation of other tests with the terminal test is difference in the functions measured by the two. The terminal SAT score represents a relatively pure measure of ability to deal with verbal concepts and relationships. Its very purity makes it somewhat narrow, and insofar as other tests sample more widely from mathematical, spatial, and other types of material, the correlations may be expected to be correspondingly reduced. The differing level of correlation for different tests probably represents in large measure differences in this degree of overlap. In particular, the lower correlations for early Stanford-Binet and early group tests may be in large measure a reflection of this difference in the function tested.

In an effort to supplement the correlations with terminal test presented in Tables I and II, correlations were also computed between pairs of earlier tests. These were computed whenever there were twenty-five cases who had taken both tests of a pair. A pair is defined both in terms of test form and in terms of the year interval in which the two tests were taken. Thus, a correlation might be computed between the American Council Psychological Examination one year prior to the Scholastic Aptitude Test and the Stanford-Binet eight years prior to the Scholastic Aptitude Test. Whenever a pair of tests defined in that way

yielded a population of twenty-five or more cases the intercorrelation was computed. In a few cases it was possible to find sets of three and even four tests for which a common population was available. In these cases all the correlations among the several tests were computed.

It was originally hoped that the correlations between pairs of earlier tests would be based upon sufficiently large groups so it would be possible to use them together with the correlations with terminal test as a basis for determining the regression weights which would yield a maximum prediction of the terminal test. Actually, the only instances in which appreciable numbers had taken the same pairs of earlier tests were those involving the American Council Psychological Examination as one of the tests. Even here, most of the groups were small and there were many gaps in the data. Furthermore, it must be remembered that the group available for a pair of tests was not identical with the group upon which the correlations with terminal test were computed. This introduces heterogeneity into the data, which makes any interpretation necessarily very tentative. Bearing these limitations in mind, it is of some interest to examine the optimum weighting of certain pairs and groups of tests for predicting terminal test performance.

Attention may be turned first to the American Council Psychological Examination. Table III shows regression weights which give the maximum prediction of the terminal Scholastic Aptitude Test. These weights have been computed for all pairs of tests for which the intercorrelations between the two tests of a pair were based upon a population of at least one hundred cases. In addition, certain weights have been computed for early Stanford-Binet testing using smaller populations, because of the special interest of these earlier tests. The first section of Table III shows weights for pairs of administrations of the American Council Psychological Examination. Examination of these weights shows that the earlier test tends in general to receive as large a weight as the later test. These results confirm the earlier observation that no decline in correlation is observed from this test as the interval is increased.

In the case of the American Council Psychological Examination certain supplementary analyses were possible. These are shown in Table IV. It was possible to get smaller groups for whom a

terminal Scholastic Aptitude Test and three earlier administrations of the American Council Examination were available.

TABLE III.—REGRESSION WEIGHTS OF PAIRS OF TESTS FOR PREDICTING CEEB SCHOLASTIC APTITUDE TEST SCORE

Later Test			Earlier Test		
Form	Year Interval	Weight	Form	Year Interval	Weight
Amer. Council	0	.39	Amer. Council	1	.39
		.33		2	.47
		.43		3	.40
	1	.30		2	.49
		.41		3	.39
	2	.50		3	.28
Amer. Council	0	.48	Otis SA, Higher	1	.30
		.51		2	.30
		.48		3	.33
		.49		4	.38
	1	.51		2	.30
		.48		3	.34
		.48		4	.37
	2	.53		3	.30
Amer. Council	0	.36	Terman Group	2	.55
		.35		4	.55
	1	.45		4	.56
		.41		5	.43
Amer. Council	0	.48	Otis SA, Intermed.	5	.35
Amer. Council	0	.47	Stanford Binet (1916)	8	.41
		.69		9	.02
		.65		10	.08
		.64		11	.20
	1	.48		8	.41
		.69		9	.02
		.62		10	.15
		.65		11	.10

Within these smaller groups the complete table of inter-correlations was computed. Regression weights were then determined from these correlations. Results for four different groupings of

American Council tests are shown in Table IV. These results again support the position that correlation between American Council and the Scholastic Aptitude tests is not a function of the interval between the two.

TABLE IV.—INTERCORRELATIONS OF SCHOLASTIC APTITUDE TEST AND SEVERAL EARLIER ADMINISTRATIONS OF AMERICAN COUNCIL PSYCHOLOGICAL EXAMINATION AND REGRESSION WEIGHTS FOR PREDICTING SCHOLASTIC APTITUDE TEST

		N	CEEB SAT	ACE 0 Years	ACE 1 Year	ACE 2 Years	Regression Weights
ACE	0 years	617	.69				.20
	1 year		.72	.82			.31
	2 years		.71	.78	.81		.30
ACE	0 years	361	.71				.29
	1 year		.72	.82			.30
	3 years		.68	.71	.75		.25
ACE	0 years	348	.69				.24
	2 years		.74	.78			.35
	3 years		.71	.72		.83	.25
ACE	1 year	363	.69				.23
	2 years		.73		.78		.35
	3 years		.71		.76	.82	.25

The next part of Table III shows data on American Council and Otis Self-Administering, Higher Examination. Here again the weight which the Otis receives shows no tendency to decrease as the interval increases. Apparently as much or more weight should be given to an Otis Test administered four years prior to the terminal test as to one administered only one year prior to the terminal test.

Most of the other data in Table III are quite fragmentary. Some attention may be given, however, to the weights for the early Stanford-Binet. The values here are quite unstable,

reflecting the instability of the basic correlations. It appears in general, however, that these early tests given at the beginning of elementary school add little to the prediction of the terminal test provided by a single administration of the American Council Examination at the secondary-school level. This should not be considered an indictment of the Stanford-Binet. Its task, as used here, is to provide an early guide as to the educational potentialities of the child. It should not be expected to supplement a test at high-school age for subsequent prediction. It should be probably considered a confirmation of the expected that this early test provides no addition to the information given by a test at the high-school level in predicting a highly verbal terminal test.

COÖPERATING SCHOOLS

Blair Academy
Brearley School
Bridgeport High Schools
The Choate School
Columbia Grammar School
Dalton School
Elizabeth High Schools
Englewood High School
Ethel Walker School
Fairfield High School
Fieldston School
Franklin School
Friends Academy (L.I.)
Glen Ridge High School
The Gunnery School
Hackensack High School
Hackley School
Hartridge School
Hopkins Grammar School
Horace Mann School (for Girls)
The Horace Mann School for Boys
Thomas Jefferson High School
The Lawrenceville School
Leonia High School
Lincoln School
The Loomis Institute
The Masters School
Millbrook School
Miss Beard's School
Montclair High School
Mount Vernon High School
Newark Academy
New Britain High School
New Rochelle High School
East Orange High School
South Orange High School

Pawling School
Pelham High School
Plainfield High School
Polytechnic Preparatory Country Day School
Pomfret School
Port Washington High School
Ridgewood High School
Riverdale Country School
Rye Country Day School
Rye High School
Scarborough School
Scarsdale High School
Stuyvesant High School
South Kent School
Taft School
Weehawken High School
West Hartford High School
Westfield High School
Westminister School
Westover School
White Plains High School
Woodmere Academy

THE EFFECT OF LANGUAGE STYLE ON READING PERFORMANCE

FRANCIS P. ROBINSON
Ohio State University

So much emphasis has been placed on vocabulary counts in books that the importance of familiarity with language style in determining reading comprehension has been neglected. Language patterns determine the relationships and the degrees of emphasis of the concepts used; facility in the recognition of these different patterns will in part determine the speed and correctness of the reader's comprehension. The purpose of this study is to show how markedly reading comprehension may be affected by complexity of language structure and something as to the factors determining skill in reading such materials.

Commonly used legal documents provide material of reader interest but of complex form. The effect of this complex language form on reading comprehension can be isolated and studied if the wording of these documents is altered to include only familiar words. The following four paragraphs represent such selections. They have been changed slightly from their originals[1] by substituting familiar synonyms for less familiar words in the text (according to Thorndike's Word Count)[2]. This rewording and the formation of the test questions were worked out under legal supervision[3]. It will be noted that the questions are straightforward and so far as possible do not involve bases for legal differences in the interpretation (one exception will be noted later). Further, the questions require comprehension of the total selection and not recognition of phrases therein.

---

[1] Sections 7, 8 Uniform Bill of Lading; Section 42.45 Wisconsin State Retirement System (and printed on the Certificate of Membership); and Cash Surrender value clause in Banker's Life Insurance Policy.

[2] E. L. Thorndike, *A Teacher's Word Book of the* 20,000 *Words Found Most Frequently and Widely in General Reading for Children and Young People.* New York: Teacher's College, Columbia Univ., 1931, pp. 182.

[3] The writer is deeply indebted to the late Dean T. Martin, Law School, Ohio State University, for assistance in preparing the materials and for administering the test to law students. Professor E. L. Bowers, Department of Economics, Ohio State University, and Professor D. B. Stuit, of the State University of Iowa, also assisted in the administration of selections to special classes.

READING TEST FOR LEGAL DOCUMENTS

This is a reading test to measure the effect upon comprehension of familiarity with a style of presentation. You are to answer the one question on each selection. Re-read the selection as often as necessary to obtain the answer. There are no catch questions and all material necessary to obtaining the answer is present in the selection.

The vocabulary has been kept as simple as possible and definitions for the most difficult words are given below:

"Bill of lading"—an account of goods shipped by any one, signed by the agent of the transportation line, thus forming a receipt for the goods.

"Liability"—an obligation for which one is liable or responsible.

"Shipper"—one who ships or sends goods.

I

"The cash surrender value shall be equal to the reserve for the benefits set out on page one of this policy, and the reserve on any existing paid-up additions and the total value of all dividends remaining at interest less any debt on the policy and a charge not exceeding one per cent of the original amount of insurance from the third to and including the ninth year, after which no surrender charge shall be made."

Question: Jones has a $10,000 insurance policy which he wishes to turn into cash at the end of the eighth year. The accumulated value of all the dividends that have remained at interest is $50. The total amount of indebtedness on the policy (including all interest payments due) is $200. The reserve set out for such a policy at the end of the eighth year is $1,300. What is the cash surrender value to Mr. Jones?

$_____

II

"The state deposit for each teacher shall be an amount equal to the following percentage upon the required deposit made by such teacher during the business year, namely to fifty per cent of the required deposit add five per cent of the required deposit for each year of teaching experience except the year for which

the deposit is made, and subtract one per cent of the required deposit for each one hundred dollars of salary received during the business year in excess of one thousand two hundred dollars, provided that the total shall not exceed two hundred per cent of the required deposit; provided, that in computing the state deposit, any amount of required deposit on account of any amount of annual salary in excess of three thousand dollars shall be omitted in applying the above percentages.

Question: Mr. Jones, a teacher, receives a salary of $3600 a year during his eleventh year of teaching on which he is required to make a 5 per cent deposit ($180) to the state teachers' retirement fund. According to the above description, what is the amount that the state contributes to this retirement fund for Mr. Jones?

$_____

III

"If this bill of lading is issued on the order of the shipper, or his agent, in exchange or in place of another bill of lading, the shipper's signature to the former bill of lading as to the statement of value or otherwise, or election of common law or bill of lading, shall be considered a part of this bill of lading as fully as if the same were written or made in or in connection with this bill of lading."

Question: A bill of lading has been copied from a signed original because the shipper's copy is worn and torn and he wants a fresh copy. The shipper must sign this new copy?

T F

IV

"The shipper shall be liable for the freight and all other lawful charges, except that if the shipper indicates by signing his name, in the space provided for that purpose on the face of the bill of lading that the railroad shall not make delivery without requiring payment of such charges and the railroad, contrary to such statements, shall make delivery without requiring such payment, the shipper (except as provided below) shall not be liable for such charges. *Provided that*, where the railroad has been instructed by the shipper to deliver said property to a receiver other than the shipper, such receiver shall not be legally liable for transportation charges in respect of the transportation of said property

which may be found to be due after the property has been delivered to him, if the receiver (a) is an agent only and does not own any part of said property, and (b) before delivery of said property has informed the delivering railroad in writing of the fact of such agency and absence of ownership, and, in the case of a shipment re-sent or diverted to a point other than that indicated in the original bill of lading, has also informed the delivering railroad in writing of the name and address of the owner of said property; and, in such cases the shipper or, in the case of a shipment so re-sent or diverted, the owner, shall be liable for such additional charges."

Question: The shipper indicates in writing on the face of the bill of lading that the shipment is not to be delivered without payment of charges by the receiver. The receiver writes to the railroad company that he does not own the shipment and that, acting as an agent, he is sending the freight to its owner at another destination. The railroad delivers the freight without collecting all of the charges at that time. Which person, persons, or company is/are responsible for all or part of the uncollected freight charges? (Answer either 'A' or 'B'. Place the number of the person or company responsible on the line or lines at the right.)

1. Original sender	A. Uncollected charges for first half of journey _____
2. First receiver	Uncollected charges for last half of journey _____
3. The owner (second receiver)	or
4. Railroad cannot collect	B. All uncollected charges for freight _____

Table I shows that these selections were written in words well within the grasp of the adult readers used in the experiment.

TABLE I.—THE VOCABULARY DIFFICULTY OF THE SELECTIONS.

Selection	I	II	III	IV
No. of words	240	92	76	133
Per cent in first 500	65	64	63	62
Per cent in first 1000	71	76	74	75
Per cent in first 5000	92	89	97	94

Further a glossary was provided for the three words not included in Thorndike's first nine thousand.

RESULTS

Different groups of college students read these selections without time limit; their results are shown in Table II. In reading selections which are brief, which deal with concrete matters and which are written in simple words, it would normally be expected that college students should have no difficulty in understanding them well enough to answer straight-forward questions of fact. It is obvious in at least three of these four selections that the use of an unfamiliar and complex style of writing in selections which are otherwise quite simple has caused difficulty with comprehension. And in passing it may be noted that if college students who represent superior readers in the total population have difficulty in reading common legal documents, then the average layman should have even more difficulty with such materials.

TABLE II.—ABILITY OF PERSONS TO READ LEGAL DOCUMENTS WHEN NOT GIVEN SPECIAL TRAINING IN SUCH READING.

	N	*Per Cent Right*
Selection I	137	40
Selection II	160	2
Selection III	258	80
Selection IV	258	52

What factors are related to facility in reading such materials? As one approach to answering this question, the students tested were divided into those above and those below the 70th percentile[4] on the Ohio State Psychological Examination. The results in Table III show that difference in intelligence has a consistent but not a statistically significant effect on ability to read involved styles of writing. Intelligence is probably a factor in reading skill for legal documents, but other factors must also be present.

Another factor may be special training in reading legal documents. The effect of such training can be shown by comparing

[4] The approximate midpoint in all of the untrained groups used.

the results of tests given to students who have had courses in insurance or in law with those of control groups matched in intelligence. The results from such an experiment are summarized in

TABLE III.—COMPARISON OF PERSONS HIGH AND LOW IN INTELLIGENCE ON THEIR COMPREHENSION FOR THESE SELECTIONS.

Selection	N in Group above	Per Cent Right		Critical Ratio[5]
		70 Percentile	Below 70 Percentile	
I	137	50	34	1.93
II	137	4	0	1.76
III	258	84	75	1.84
IV	258	63	51	2.00

[5] In terms of the standard error of a proportion.

TABLE IV.—COMPARISON OF TRAINED AND UNTRAINED GROUPS (MATCHED FOR INTELLIGENCE) IN THEIR ACCURACY OF COMPREHENSION.

Selection	Per Cent Right		Critical Ratio
	Trained	Untrained	
I	82	40	6.10
II	31	2	3.35
III	89	81	1.95
IV (total)	58	52	0.94
Ans: '3'	28	35	1.21
Ans: '1' and '3'	30	17	2.28

Table IV. About half of the seventy-eight insurance students had had one course on insurance; the rest, two such courses. The seventy-eight law students consisted of freshmen and seniors from two different law schools.

Since professional groups tend to be brighter on the average than the normal run of college students, an untrained group,

equal in ability to the trained students, had to be selected. The intelligence quartiles for the insurance group and its control were 91, 75, and 33 percentiles, and for the law group and its control were 93, 79, and 63 percentiles. Thus equated, one is able to say that any differences found are not due to intelligence but are due to training (or to other factors not controlled).

It will be noted in Table IV that the differences in favor of the group trained in reading insurance (selections I and II) are statistically significant. Reading skill with such material is more than knowledge of vocabulary and intelligence; specific training should be given in handling this language form. While the differences on selections III and IV are in favor of the trained group, these differences are not statistically significant. These smaller differences may be due to various causes: (1) Selections III and IV may have been so easy that special skill did not have much chance to operate. (2) There may have been some difference in motivation among the trained groups. Selections I and II represented parts of final examinations while selections III and IV were done as extra outside assignments which did not count on grades. And (3) there may be more transfer from reading psychological literature to typical legal writing than there is to insurance material. A related fact is the presence of mathematical computations, albeit quite simple ones, in the first two selections.

The difference in the preference by the trained and the untrained for one of the two correct answers to Selection IV is interesting. Both '3' and '1 and 3' are correct answers depending in part on one's interpretation of the word 'additional' in the last line, i.e. does it refer to the charges after the freight is reconsigned or to charges in addition to those already paid? While the difference in preference for one of these two answers is not statistically significant, it may be indicative of a difference in interpretation due to reference to different patterns of thinking.

CONCLUSIONS

The results of this study illustrate the following points concerning the psychology of reading.

1) Vocabulary is not the only determinant of paragraph comprehension; language structure is important.

2) Indications were found that both intelligence and training seem to affect facility in reading a complex language style.

3) There seems to be sufficient evidence to indicate: (a) That legal style should be simplified for the average layman's use or that laymen should be taught to read such material. The insurance and the retirement act were outstanding examples of this. (b) That in insurance courses, and possibly in law courses, some emphasis must be placed on training in 'how to read' such materials.

THE ABILITY OF CHILDREN TO READ CROSS-SECTIONS[1]

MORTON MALTER

Research Fellow, Department of Education, The University of Chicago

Reading materials for elementary-school children ofttimes contain a number of diagrams. According to Hoban et al.[2] the diagram is "perhaps the most abstract of illustrative material. . . . While it may use some pictorial means it usually reduces facts and ideas to the strictest convention of line and geometric form." These authors then state—and most educators would agree—that "The most effective use of the diagram comes after the pupil has been shown the original fact" in a more concrete form.

At times, diagrams depict cross-sectional views: in those instances we refer to the diagram as a 'cross-section.' A cross-sectional view can be shown in a photograph or realistic drawing. Photographs are seldom employed for this purpose because they do not reproduce the intrinsic detail with sufficient clarity. The difference between a realistic drawing and a diagram is essentially the extent of abstraction involved; in fact, it is sometimes difficult to classify an illustration as being of one type or the other. It would be difficult to say, for example, whether the doll in Fig. 2 were realistically or diagrammatically drawn. In general it is useful to consider most cross-sections as being a type of diagrams.

Cross-sectional views can be prepared by one of two methods, and, consequently, there are two types of cross-sections. Fig. 1 shows the procedure for preparing cross-sectional views by Methods 1 and 2. It is not possible to tell which of the methods was employed in preparing the cross-sectional views that resulted in A-E in Fig. 2. In this particular case the employment of

---

[1] Hoban, Charles F., Charles F. Hoban Jr. and Samuel B. Zisman, *Visualizing the Curriculum*, New York, Dryden Press, 1946, p. 249.

[2] This is one of a series of studies being conducted in the University of Chicago Center for the Study of Audio-visual Instructional Materials. This investigation was subsidized in part by a grant from the Enclyclopedia Britannica Jr. awarded by its Advisory Committee.

either method would have resulted in these kinds of cross-sections. The cross-sections A to E in Fig. 3, however, definitely

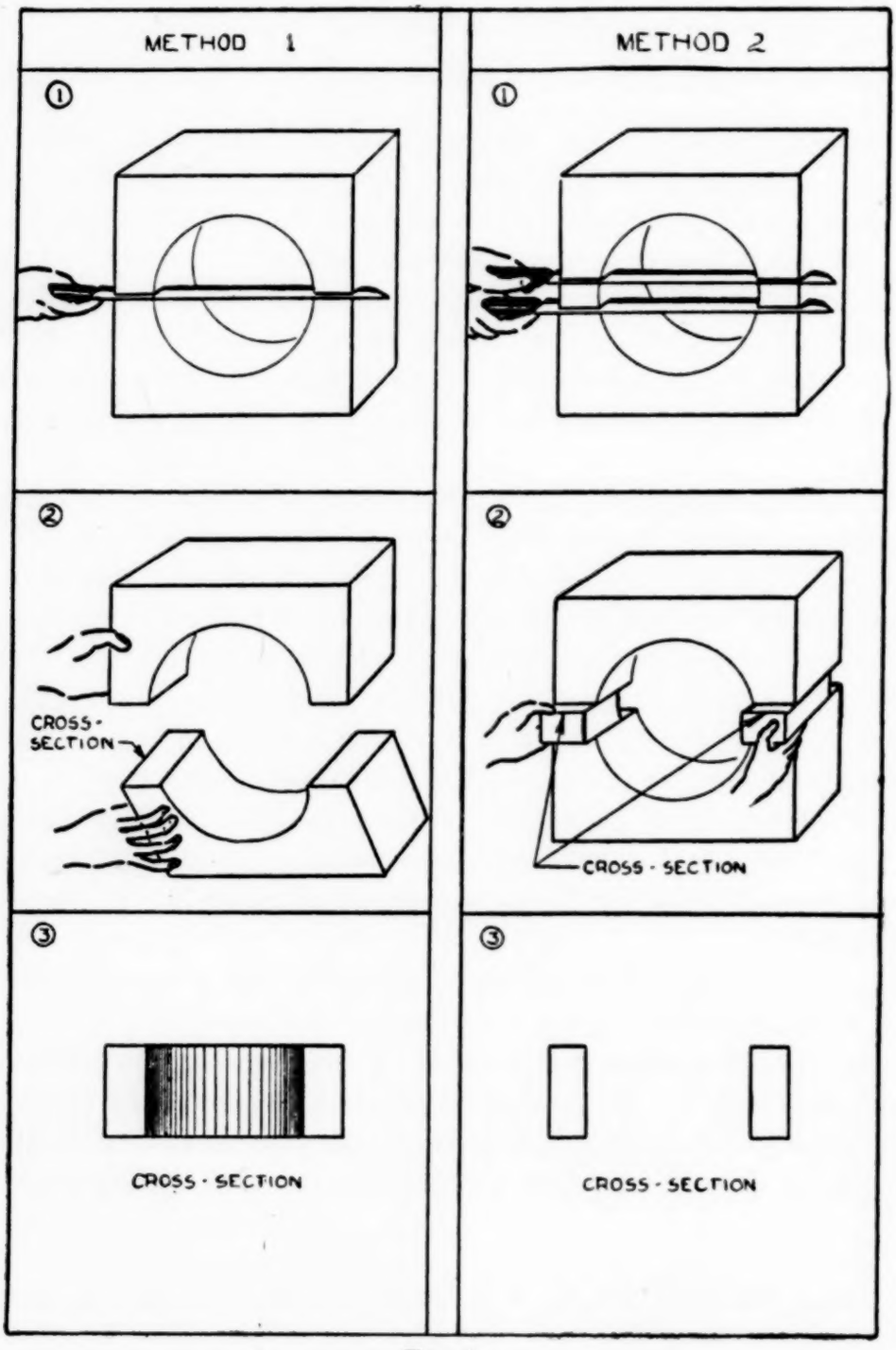

FIG. 1.

represent views prepared by Method 2. This is indicated by B. If B in Fig. 3 were not a cross-sectional view prepared by Method

2, the white space between the two halves would be occupied by some kind of hatch-marks indicating a concave recess.

The writer has not found any reports of the ability of elementary-school children to read cross-sections. Since children do meet such diagrams, oft times in the initial stages of their learning, information about their ability to read them seems essential.

PROBLEMS

The hypothesis that the writer wished to test was that elementary-school children, grades IV to VIII, cannot read cross-sections. The term 'read,' when used in reference to cross-sections, can mean at least two things. In a special sense, it refers to the act of examining a cross-section and then mentally transposing it to its place of origin. For example, the child who is aware that E represents the view exposed by Cut 1 in Fig. 2 is in a special sense reading a cross-section. If E happened to appear alone, the child, of course, would have to extrapolate the doll.

Reading a cross-section in a broader sense involves two interrelated steps. The first in reading in the special sense that was just described. Secondly, reading a cross-section involves going on and interpreting or deriving meaning from the parts contained within the cross-section. Reading a cross-section in the broader sense might be thought of as reading it at a higher level of understanding.

In this study the term 'read,' when applied to cross-sections, is used in the special rather than the broader sense.

PROCEDURE AND RESULTS

Two illustrations (Figs. 2,3) were prepared. The central object in Fig. 2 is a doll; in Fig. 3 it is a 'stand.' The illustrations were planagraphed and presented as part of a test battery to three hundred fifty-eight children, grades IV to VIII, in three schools, A, B and C, located in Wisconsin, Indiana and Illinois, respectively. Schools A and B were public institutions; School C was a private one. The children in School C were superior to those in Schools A and B in respect to intelligence. (See Table 4.)

Answer sheets were given to the children, along with the following directions, presented in both oral and written form:

"The figure underlined in red has been cut across in four places and it is surrounded by five cross-sections. There is one proper

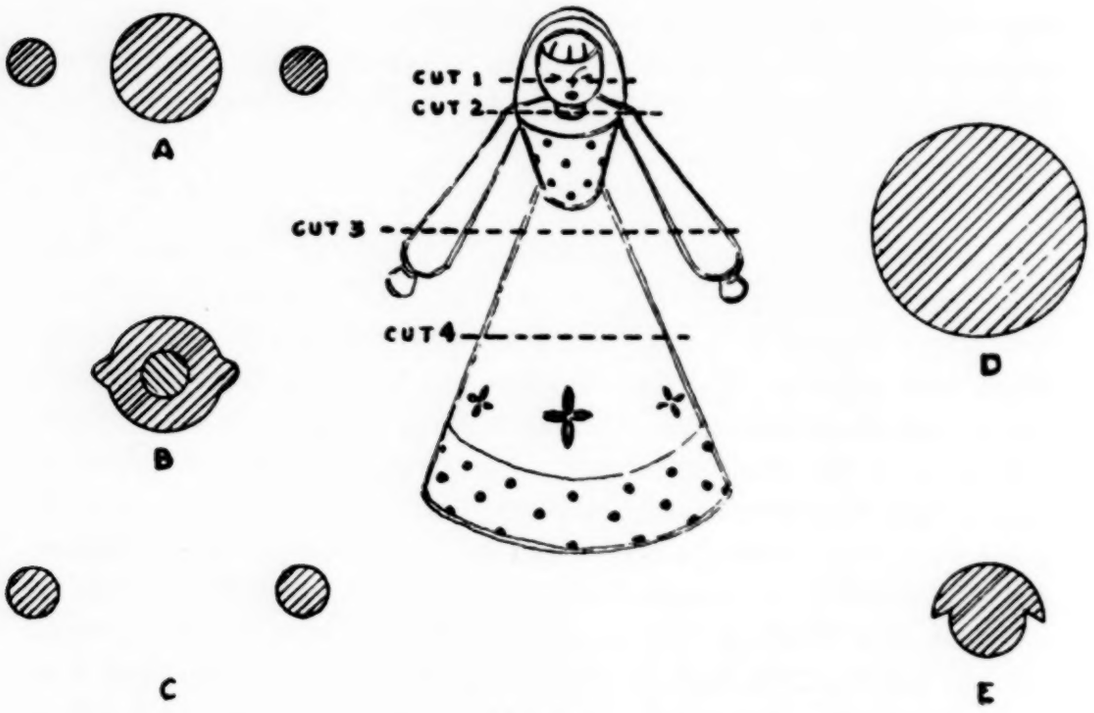

FIG. 2.

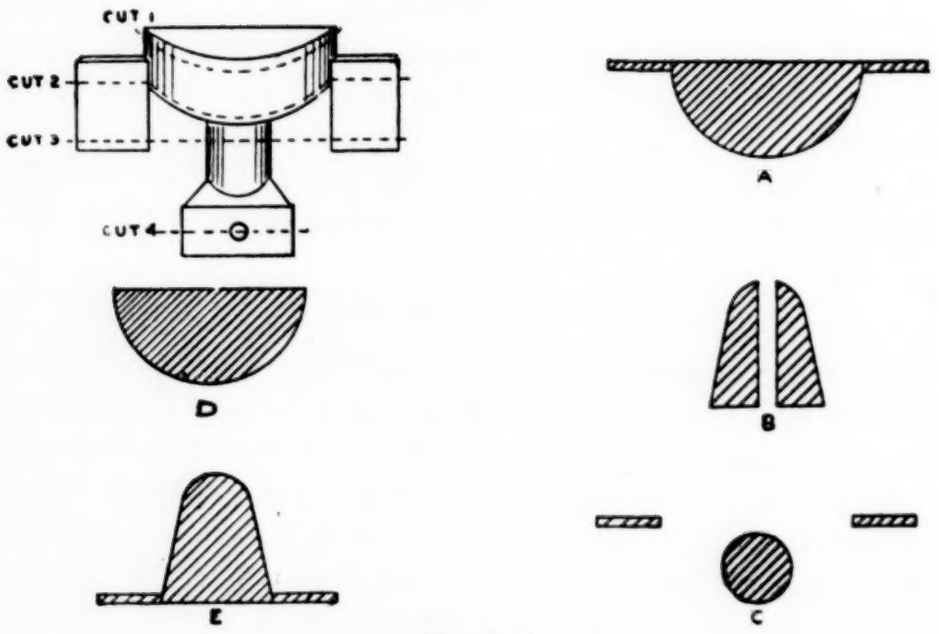

FIG. 3.

cross-section for each cut. Show that you know what cross-section is made by each cut by matching them up. You can do this by placing the proper letter after the cut number. (on the

answer sheet.) There will be one cross-section left over that does not belong to any cut."

The writer believes, from speaking with the children after the test, that they understood the directions.

The test consisted of eight items. Each paper was scored right or wrong on each of the items, and, consequently, a score of 8 was perfect. No corrections were made for guessing.

Since the main hypothesis being tested was that elementary-school children, grades IV to VIII, cannot read cross-sections, the following statistical technique was employed:

TABLE I.—OBSERVED AND THEORETICAL DISTRIBUTION OF SCORES FOR ALL OF THE STUDENTS TESTED IN SCHOOL A

	Test Scores								
	8	7	6	5	4	3	2	1	0
Observed distribution...	6	8	8	17	21	18	19	30	34
Theoretical distribution.	0	0	0	1.47	7.25	23.51	46.69	53.94	26.88

First for school A: The point binomial $(p + q)^n$ was expanded: $p = \frac{1}{5}$: $q = \frac{4}{5}$; $n = 8$. The expansion of the binomial resulted in a series of decimals, which summed to unity. Each of the decimals was then multiplied by 161, which represented the total population in School A. This resulted in a theoretical distribution of all of the one hundred sixty-one cases, assuming that nobody could read cross-sections and that chance alone had been operating during the test. Table 1 shows this theoretical distribution: thus, no cases would have been expected to receive scores of 8, 7 or 6: about one case a score of 5: seven cases a score of 4; and so forth.

Next, the observed scores made by all of the one hundred sixty-one students were distributed according to number correct. Table 1 shows this distribution; thus six cases received a score of 8 on the test; eight a score of 7; and so forth.

It was next planned to apply the chi-square formula and see if the difference between the theoretical and observed distributions

was significant. However, the mere inspection of Table 1 made this step unnecessary; the disparity between the theoretical and observed differences was apparent at once. For example, no cases would have received a score of 8 if chance alone determined the score; actually six cases did receive this score. Or, to take another example, 1.47 cases would have received a score of 5; actually seventeen students did receive this score. Obviously the hypothesis that elementary-school children (those in School A) cannot read cross-sections can be rejected.

TABLE II.—OBSERVED AND THEORETICAL DISTRIBUTION OF SCORES FOR ALL OF THE STUDENTS TESTED IN SCHOOL B

	Test Scores								
	8	7	6	5	4	3	2	1	0
Observed distribution...	7	11	4	7	13	15	10	10	18
Theoretical distribution.	0	0	0	.87	4.27	13.87	27.55	31.82	15.86

TABLE III.—OBSERVED AND THEORETICAL DISTRIBUTION OF SCORES FOR ALL OF THE STUDENTS TESTED IN SCHOOL C

	Test Scores								
	8	7	6	5	4	3	2	1	0
Observed distribution..	17	22	8	13	9	10	9	4	10
Theoretical distribution	0	0	0	.94	4.59	14.89	29.58	34.17	17.03

The same statistical technique was used in the cases for Schools B and C. For these schools, however, the decimals resulting from the expansion of the binomial $(p + q)^n$ were multiplied by 95 and 102, the total populations in Schools B and C, respectively. The theoretical and observed distributions for Schools B and C are shown in Tables 2 and 3. Again, on the basis of the results shown in these tables, we can reject the hypothesis that children (those in Schools B and C) cannot read cross-sections.

Table 4 shows the mean scores made by each grade tested in all three schools. In School A and B[3] (normal populations) the mean score increased progressively from grades IV to VIII. In School C this progressive trend was upset by the results obtained in Grade V, which were better than those obtained for grades VI and VII. The better score made by the children in Grade V might in part be accounted for by the fact that they averaged six points higher in intelligence than the children in Grades VI and

TABLE IV.—MEAN SCORES MADE IN GRADES IV TO VIII IN SCHOOLS A, B, AND C

	No. Tested	M CA	M IQ	SD IQ
School A.....	161	12.1	101	11.40
School B.....	95	12.28	100	13.25
School C.....	102	11.73	131	11.67

Grade	Mean Score		
	School A	School B	School C
4	1.38	1.33	2.63
5	2.12	2.50	4.95
6	2.80	3.80	3.83
7	3.51	4.24	4.94
8	3.87	4.53	6.64

VII. The fact that the children in School C did better on the whole than those in Schools A and B is not surprising. They are exceptionally bright children and consistently rank one or more grades above the norm on most standardized tests.

---

[3] Since the children in Schools A and B represented similar elementary-school populations, it was decided to see if the observed test distributions in these schools differed significantly. The exact hypothesis that we wished to test, then, was that there was no relationship between school attended and performance on this test. Tables 1 and 2 show the observed distributions for these two schools. The value of chi-square was then computed and it equaled 7.52. For 6d.f. (degrees of freedom) this value of chi-square is exceeded between twenty and thirty per cent of the time. We may, therefore, accept the hypothesis stated above.

INTERPRETATION OF RESULTS

Because the test was part of a diagrammatic reading battery it was made short. The reliability of the results are thus questionable. Furthermore, although the test was given to three hundred fifty-eight children, they were selected from only three schools.

On the other hand, the following might be said in favor of the results:

a) After the test was administered the writer spoke to the children in the various classes. Many of them frankly admitted that they did not know how to read cross-sections and were often puzzled by them.

b) The results obtained for Schools A and B (normal population) were similar. As indicated in footnote 3, the difference between the results obtained in the two schools was not statistically significant.

The hypothesis that the children tested cannot read cross-section was rejected. This does not imply, however, that all of the children tested can read cross-sections. As Tables 1–3 indicate, some of the children could read all of the eight sections, others could read some of them, while still others could read none. The rejection of the main hypothesis simply means that the writer was wrong in assuming that none of the children in grades IV to VIII would be able to read any of the cross-sections in the test.

As far as the writer could determine, none of the children ever had received any special training in reading cross-sections. Yet, as Table 4 indicated, the mean scores increased progressively through grades IV to VIII in Schools A and B. (normal populations). However, the following should be noted: the highest mean score made in Schools A and B were only 3.87 and 3.28, respectively. Since chance alone would account for mean scores of 1.6, scores of 3.87 and 3.28 cannot be considered significant.

Finally, only 8.4 per cent of all of the children tested received perfect scores of 8; thirty-eight per cent received scores of 1 or 0. This leaves 53.6 per cent of the children with scores that are better than can be accounted for by chance alone, but yet not perfect. Because of the brevity and design of the study, no attempt was made to account for these scores.

CONCLUSION

The scores made by most of the children were better than could be accounted for by chance alone. However, the scores indicate that most of the children have not developed to a place where they can read any cross-section. It seems fair to conclude that most of the children are unable to read many of the cross-sections that appear in their reading materials.

IMPLICATIONS

Studies of this kind are practically useful only in the extent to which they result in effective action. The fact that there is a reading problem connected with pictorial materials is frequently overlooked.

If the need arises, children can be taught how to read cross-sections in Grade IV, perhaps sooner. This statement is made on the basis of a study made in Grade IV of School A. The children in that class expressed an interest in learning how to read the cross-sections shown to them during the test. The writer spent about ten minutes teaching this skill, using Fig. 2 in his explanation. The children were then shown Fig. 3 and asked to perform again as they had during the test. The first time that the children performed on the four items in Fig. 3 they made a mean score of .79. After being taught for ten minutes the mean score made was 2.73. One month later the children were again examined with Fig. 3. This time they made a mean score of 2.56. The experiment was uncontrolled, but the writer believes that the teaching accounted for the increase in score.

The contention might be advanced that teaching children to read cross-sections in the special sense, as previously mentioned, is not sufficient. The argument might run as follows: Reading a cross-section in the special sense is a means to an end. The end is reading the cross-section in the broader sense or deriving meaning from it. Unless the precipient already has seen the cross-sectional view in a more concrete form, there is a danger that the diagram will not have meaning for him even if he can read it in the special sense. The argument might conclude; if the percipient has experienced the cross-sectional view in a more concrete form, he should have no trouble reading it in the special sense.

Implicit in the above argument is the belief that cross-sections are meaningful to children only if based on a more concrete foundation (i.e. experiencing the actual item, models, motion-pictures). The writer considers this viewpoint reasonable, but there is little, if any, evidence to support it. One reason why cross-sections are not meaningful to children might be due to the fact that they have never learned how to read them in the special sense.

A TEST TO MEASURE APTITUDE IN THE HEBREW LANGUAGE

NOAH NARDI

Psychologist, Jewish Education Committee of New York

THE NEED FOR TESTING HEBREW APTITUDE

The ability of a pupil to master the Hebrew language is a prerequisite to whatever success he may have in Hebrew school. It determines the length of his Jewish education and affects his future attitude toward Jewish values and ways of life. Important as the child's aptitude for the study of Hebrew is, it is very much taken for granted and is often confused with mental testing and IQ ratings.

Aptitude in general is subject to influences and variations according to the situation involved. In the study of Hebrew, factors of age, attitude, quality of instruction, etc. will of course greatly affect the pupil's chances of success. However, while mental tests were found helpful in predicting general school achievement, it has been equally proven that such tests cannot be depended upon for predicting achievement in specific subjects, especially in the study of foreign languages.

Ross and Hooks examined more than three hundred coefficients of correlations gathered from studies concerned with the relationship between intelligence and achievement in various high-school subjects and calculated the following medians:

IQ's and school marks in Science	.44
IQ's and school marks in Mathematics	.37
IQ's and school marks in Foreign Languages	.33
IQ's and school marks in History and English	.45[1]

Thus the correlations between mental tests and achievement in the foreign languages are lower than those between intelligence tests and the achievement in general high-school subjects. This is true of foreign language studies on the college level as well. One of the best available summaries of data bearing on this problem is that compiled by Segel.[2] His data of correlations between

[1] "Prognosis" in *Encyclopedia of Educational Research*. New York: The Macmillan Co., 1941, p. 846.

[2] Segel, David. *Prediction of Success in College*, U.S. Office of Education, Bulletin #15, 1934.

scholarship in foreign languages and mental and aptitude tests on the college level are summarized below:

Foreign Language	General Mental Tests	Aptitude Tests
French	.32	.47
German	.36	
Latin	.41	
Spanish	.23	

Further data concerning the prediction of success in foreign languages at the college level were given by Tallent[3], who reported the following correlations:

Modern Language marks and IQ	.21
Modern Language marks and English placement scores	.49
Modern Language marks and English marks	.56

These and other findings lead to the conclusion that for predicting achievement in foreign languages, tests of language aptitude are superior to those of general intelligence. Even so, the relationships are not sufficiently high to assure the necessary degree of accuracy in making predictions regarding individual students.

As to prognosis on the elementary-school level, foreign languages are not taught in the elementary school, and no aptitude tests have been devised for this age level. The relationship between IQ scores and elementary-school subject was summarized by Louttit and is as follows:

Mean correlation of studies in reading	.55
in spelling	.49
in arithmetic	.55
in handwriting	.13[4]

These correlations reveal a considerable prognosis for general mental tests in relation to reading, spelling and arithmetic. Whether or not this prediction will hold true for achievement in Hebrew is another matter. It stands to reason that it will not,

---

[3] Tallent, Ruth E. "Three Coefficients of Correlations that Concern Modern Foreign Languages," *Modern Language Journal*, 22: 591–94, 1938.

[4] Louttit, C. M. *Clinical Psychology*, Harper, 1936.

because Hebrew, neither in form nor in character, has much affinity to the academic elementary school-subject.

In summing up the results of research on prognosis in the general field at all levels—elementary school, high school and college—it becomes evident that the predictive value of the different measures used is somewhere between .40 and .60. In terms of percentages, this predictive value means from thirty-five to forty-three per cent better than chance. The possibility of making use of already available instruments of prediction from the general education field in the Hebrew schools is, therefore, rather limited. Hence the need for Jewish aptitude tests, especially in Hebrew—the pivotal factor in the success of Jewish schooling.

PREVIOUS ATTEMPTS

Other attempts in devising prognosis tests in Hebrew are those by Evelyn Garfiel and Julius B. Maller. Garfiel experimented with a number of study-test lessons on a rather limited sampling of pupils. The report of this study[5] bears the distinction of being one of the first articles in statistical and educational research in Jewish education. However, no definite test of Hebrew prognosis evolved out of this pioneer attempt.

Such a test was prepared in 1929 by Maller.[6] It was intended originally for high-school and teacher-training students, with only two of its six subtests specifically designed for the testing of Hebrew learning ability. An attempt to adjust this test to the elementary-school level was made a year later by Simon Silverman.[7] The latter test, however, was not sufficiently validated. It is heavily weighted by an English vocabulary test which makes up fifty of the total two hundred twenty items, and accounts neither for the validity of the individual items nor for the relative weight of each of the subjects. It is expressed in raw scores only and does not provide for reliable and representative norms.

---

[5] "Hebrew Prognosis Test," Evelyn Garfiel, *The American Journal of Semitic Languages and Literature*, Vol. XLV, 1, Oct., 1928.

[6] "A Standardized Hebrew Prognosis Test," Julius B. Maller, *Jewish Education*, Vol. III, 1931.

[7] *Hebrew Prognosis Test for Elementary School Grades.* Julius B. Maller, and Simon S. Silverman. Department of Synagogue and School Extension, Union ofAmerican Hebrew Congregations, 1931.

8. ***Thin branch*** in Hebrew is עָנָף דַּק. ***Thin*** is דַּק in Hebrew. What is ***branch***?

1. עוֹף 2. כָּנָף 3. אַף 4. עָנָף 5. עָנָן

9. ***It is cold in the winter*** in Hebrew is קַר בַּחֹרֶף. ***In the winter*** is בַּחֹרֶף. How do you say ***It is cold*** in Hebrew?

1. קַר 2. כַּר 3. קַשׁ 4. קֵן 5. רַק

10. ***Jacob looked in the mirror*** in Hebrew is יַעֲקֹב הִסְתַּכֵּל בָּרְאִי. ***In the mirror*** is בָּרְאִי. ***Jacob goes*** is יַעֲקֹב הוֹלֵךְ. How do you say ***looked*** in Hebrew?

1. הִתְהַלֵּךְ 2. הִסְתַּכֵּל 3. הִסְתַּדֵּר 4. מִסְתַּכֵּל 5. הִתְחַכֵּם

Number Right_____
Standard Score_____

TEST V — MEMORY

Below are two rows of words, one in Hebrew and one in English. Study them carefully and try to remember as many as you can.

Came	בָּאָה	Sabbath	שַׁבָּת
Day	יוֹם	Rested	שָׁבַת
Joy	שִׂמְחָה	Rest	מְנוּחָה
God	אֱלֹהִים	Welcome	שָׁלוֹם

Write the English meaning of each Hebrew word.

1. בָּאָה .....
2. יוֹם .....
3. שִׂמְחָה .....
4. שַׁבָּת .....
5. שָׁבַת .....
6. מְנוּחָה .....
7. שָׁלוֹם .....
8. אֱלֹהִים .....

Before turning to the next page, study these words so that you will be sure to know them from memory for the next question.

FIG. 1. Specimen page from Hebrew Aptitude Test.

PURPOSE OF THE TEST

The Hebrew Aptitude Test is designed to predict the degree of success that a pupil, other things being equal, will have in learning Hebrew.[1] It is intended for the age level of seven to seventeen and is to be used by pupils with no Hebrew knowledge whatsoever. The test will be found helpful in the classification of beginners' classes into rapid and slow progress groups, and can also be used as a means of discovering the lazy but able pupils. While the test (as all other aptitude examinations) may not have high accuracy in indivdual prediction for pupils of medium ability, it will be found useful, nevertheless, in singling out the poorest and the ablest pupils.

NATURE OF THE TEST

Subtest I—Recognition of Hebrew characters in single letters, syllables and whole words. Twelve items to be matched out of twenty-eight choices.

Subtest II—Form Comprehension. Consists of simple words and elementary rules of the Hebrew language. The Hebrew words are in English letters. Twenty items.

Subtest III—Translation. A short story in Hebrew written in English letters with a word list provided for each word of the story. The test consists of filling in missing words of a comprehensive reading nature. Twelve items.

Subtest IV—Identification. Recognition of individual Hebrew words in phrases and sentences through the process of elimination. Ten items.

Subtest V—Memory. Exercise and retention of Hebrew words and phrases. Twenty items.

RELIABILITY

The reliability of the test has been determined by the split-half method. The coefficient for the two halves was found to be .76, and, when adjusted by the Spearman-Brown formula, the corrected value for the full length of the test was found to be .89, which is satisfactory for a group aptitude test.

---

* A specimen page from the test (reduced by about one-third) is presented herewith. The complete test is distributed by Jewish Education Committee of New York, 1776 Broadway, New York City.

VALIDITY

In establishing the validity of the test, two criteria were used: (1) The internal criterion; namely, validation of each individual

LIST OF VALIDITY CORRELATIONS

No.	Variables	Type of Group	No. of Cases	Correlation
1.	Aptitude and IQ ratings	Public High School	54	.56
2.	Aptitude and IQ ratings	Hebrew School Group	77	.31
3.	Aptitude and teachers' marks after the first two weeks of Hebrew study	Combined Group	95	.30
4.	Aptitude and teachers' marks of pupils who left school after one term of study	Hebrew School Group	27	.63*
5.	Aptitude and Hebrew marks after two more terms of Hebrew study	Public Junior High School	20	.83*
6.	Aptitude and Hebrew marks after three and more terms	Hebrew School Group I	58	.56*
7.	Aptitude and Hebrew marks after three terms of Hebrew study	Hebrew School Group II	29	.75*
8.	Aptitude and Hebrew marks after four terms of Hebrew study	Hebrew School Group	13	.88*
9.	Aptitude and Hebrew marks after five terms of Hebrew study	Hebrew School Group	16	.74*
10.	Aptitude scores and teachers' marks	Hebrew School Group	17	.64 (Rho)

* Correlation non-linear. This is to be expected in aptitude testing, where the aptitude test is expressed in quantitative scores and the teachers' marks are of a discrete nature.

item and inter-correlation of each of the subtests and (2) the external criterion of teachers' evaluation of pupils' progress in Hebrew. Comparisons were also made between aptitude and IQ scores, and between IQ scores and teachers' marks.

The correlations between aptitude scores and teachers' marks were quite satisfactory. These correlations should be viewed, however, with reservation because of the doubtful reliability of teachers' marks. Correlations were calculated after one to four terms of Hebrew study, but one must keep in mind that achievement is determined not only by aptitude, but also by many other factors such as attitude, home environment, effectiveness of instruction, studiousness, outside activities, etc. One can, therefore, reasonably assume that the Hebrew Aptitude Test is probably more highly predictive of ability in Hebrew than the correlations reported would indicate. This can be readily seen from the fact that the Hebrew Aptitude scores showed higher prediction in achievement of Hebrew than that obtained by intelligence tests.

ITEM VALIDATION

The work or effectiveness of a test depends very much on the extent to which each of its individual items contributes to the validity of the whole test and upon its power to discriminate between pupils of high and low aptitude. Different statistical techniques have been developed to express the validity of the individual item in terms of a single numerical value. The technique used for the Hebrew Aptitude Test was that of bi-serial correlations. Out of all bi-serial r's derived for each of the items in the test, only four were below .30 and in need of revision. The majority of the items, however, were from .60 to .85 as indicated in Table I.

SUBTEST INTERCORRELATIONS

Item validation is but one of the internal criteria in the establishment of test validity. Another important internal criterion is the extent to which the subtests correlate with each other. The assumption is that the correlation between the subtests themselves should be as low as possible, while the correlation between the tests and outside criteria should be as high as possible. The reason for this is that insofar as subtests correlate

with each other they are testing the same functions, and the higher this correlation is, the stronger the probability that both subtests measure the same function and, therefore, duplicate each other.

TABLE I.—DISTRIBUTION OF BI-SERIAL CORRELATIONS OF THE HEBREW APTITUDE TEST

Step Interval	Frequency
.80–.84	2
.75–.79	6
.70–.74	17
.65–.69	8
.60–.64	9
.55–.59	8
.50–.54	1
.45–.49	3
.40–.44	3
.35–.39	3
.30–.34	3
.25–.29	3
.20–.24	1
Total	67

Table II gives the subtest intercorrelations for the Hebrew Aptitude Test. They vary from .16 to .55 and are considerably lower than the correlation of the total test with the criterion.

TABLE II.—INTERCORRELATION OF SUBTEST HEBREW APTITUDE TEST

	II	III	IV	V
I	.35	.16	.24	.23
II		.47	.45	.34
III			.55	.27
IV				.23

STANDARD SCORES

The scores on the Hebrew Aptitude Test are easily converted into standard scores. These scores were calculated according to

chronological age and are based on samplings tested in 1942 and 1943.

The mean scores of each of the subtests are presented graphically according to ages seven to seventeen in Chart I. It proves

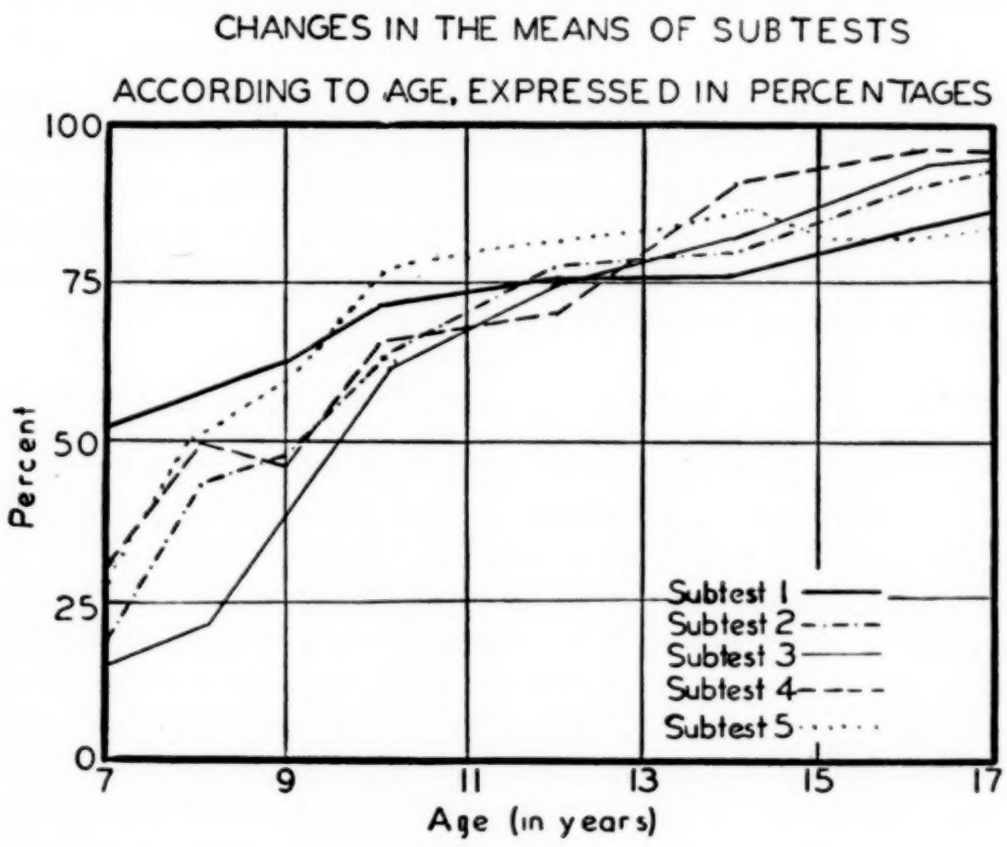

CHART I

that the age variable seems to constitute a definite factor in the Hebrew Aptitude scores.

NORMS

The sampling of about three hundred pupils used for the Hebrew Aptitude Test in the first two editions of the test is fairly representative of the group for which the norms are intended. All pupils tested were beginners in Hebrew with no or very little previous experience in the Hebrew language. The norms are derived on the basis of age. Most of the pupils tested were from Hebrew afternoon schools and were in the seven- to twelve-age range. Those of thirteen to seventeen were pupils of Hebrew in public junior and senior high schools. The test is applicable to higher ages as well.

In order to interpret more conveniently the standard score norms, the procedure of assigning the conventional letter grades to standard scores has been employed here. Each age group was divided into five subgroups: A, B, C, D and F, with the range of ability equal in each group to 1.2 of its standard deviation.

TIMING

The time needed for answering each of the five subtests has been experimentally determined. The total working time for the whole test is forty-eight minutes. At least fifteen more minutes should be allowed as time is consumed in reading the directions and distributing and collecting materials.

ADMINISTRATIVE FEATURES OF THE TEST

The method of administering and scoring the test is given in the "Directions for Administering and Scoring of Hebrew Aptitude Test" provided with the test. Specific directions to be followed exactly are given for administering each of the five subtests. The scoring of the test is done mechanically through scoring keys. The number of questions on a test answered right is the raw score for that test. Corresponding to each raw score of the subtest is the standard score given in the Manual. The composite score for all of the five subtests is the average standard score of these subtests. Letter ratings for the test are provided in the Manual and the standard score for each subtest or the composite score of the whole test can be easily transmuted into a corresponding letter rating.

THE EFFECT OF PERSONALITY ON LANGUAGE ACHIEVEMENT

HAROLD B. DUNKEL

Board of Examinations, University of Chicago

The prognosis of achievement in foreign language courses has never been easy. The L-scores of the ACE Psychological Examination and similar measures of verbal intelligence show fairly high positive correlation with success in elementary language classes; but the coefficients are always sufficiently low to demonstrate clearly that other factors are involved,[1] and the language aptitude tests have furnished little, if any, better prediction. Yet the isolation and estimation of these other forces influencing language achievement would be valuable to language teachers in a variety of uses ranging from the general theory of language learning to the practical counseling of students.

Many suggestions have been offered from time to time to account for this discrepancy between verbal intelligence and linguistic performance—interest and motivation, study habits, personality traits, and the like. Some data available at the University of Chicago in 1946–47 offered an unusual opportunity to explore one of these possibilities—personality type. The Rorschach test was given, by group administration, to about one hundred fifty students entering the College at the eleventh grade;[2] and most of these students also took the full battery of entrance-tests, which included the ACE Psychological and a placement test in one of four foreign languages. Of these four language groups, the largest for which complete data were available was that composed of those forty-six students who had had two years of Latin in high school and who took that placement

---

[1] For a recently published study with rather typical findings, see A. G. Bovée and G. J. Froelich, "Some Observations on the Relationship between Mental Ability and Achievement in French," *Modern Language Journal* XXX (October, 1946) pp. 333–6. They report, for a series of sixteen coefficients, a median coefficient of .52 between the Stanford-Binet (1937 revision, Form L) and the Coöperative French Test (various forms).

[2] Somewhat smaller groups were also available at the three next higher grades, but they involved variables (such as the interval between originally studying the language and taking the placement test) which made them much less useful for the present study.

test. The correlation between the L-score of the Psychological test and the Latin placement was relatively low, .21.

The best method of examining the data from the standpoint of prediction seemed to be to begin by advancing hypotheses as to which personality types were most likely to succeed and to fail in the first two years of Latin and then to test these blind hypotheses on the data. For reasons which need not be developed in detail here, it seemed most likely that what may be roughly identified as a mildly 'compulsive' personality would be that with the greatest probability of success in elementary Latin —at least as it is commonly taught. The meticulous student who gives close attention to details, who keeps a relatively tight rein on his imagination and impulses, and who is somewhat introversive, appeared likely to have a better chance of success. Conversely, the more impulsive, happy-go-lucky, or phantasy-making personalities would seem likely to have found elementary Latin less congenial and, hence, to have been less successful in it than their basic verbal ability would make possible.

Twelve students whose Rorschach patterns indicated the somewhat rigid and meticulous make-up were found in the group.[3] The effort to find a contrasting 'impulsive' group was much less successful since this type was not well represented in the group available.[4] Consequently, the twelve students selected were simply paired with other members of the total group on the basis of the best match in psychological scores, especially the L.[5] Because of the fairly wide choice, pairing was easy and

---

[3] Elements in the general sort of Rorschach configuration sought were: high F%, high Dd and/or S%, M much greater than C with little or no C, FC > CF, the ratio (H + A):(Hd + Ad) favoring the details, and similar points. (The symbols and general interpretation of the Rorschach results follow B. Klopfer and D. M. Kelley, *The Rorschach Technique*, World Book Co., 1942.)

I am under great obligation to Mrs. Harriet Moore, of the University Clinics, who scored all the Rorschach tests and whose interpretations I could use as a check on my own.

[4] It should be pointed out that the group administration of the test may have some influence on the results, at least, on the type of configuration sought here. Seeing the slide projected at some distance rather than holding the card in the hand may cause the student to give an unusually high W-response; as a result, the scoring of the location categories (W, D, d, and Dd) and of the order of dealing with these parts (Suc.) is slightly difficult.

[5] The Board of Examination keeps its records in terms of standard scores

rather exact. The combined L-scores for the experimental group of twelve totaled 267; that of their pairs, 270. On the Latin placement, however, the experimental group failed to show any superiority, with the total scores for the two groups being 262 and 263, respectively. One student in what I am loosely calling 'the compulsive group' did make a much higher score than would have been anticipated on the basis of his L-score. But others in the experimental group did not; and some members of the control group, of various other personality types, made equally large gains. Thus the original hypothesis, that this particular kind of personality was a major element in the student's success, was not substantiated.

An obvious next step was to examine the actual cases which exceeded or fell short of the prediction which could have been based on their L-scores and to see what Rorschach patterns actually appeared in these two groups. Once more, twelve students fell into the group which exceeded expectations by three or more standard-score points, and ten into the group who fell short by a similar margin. In the consideration of these groups, one factor inherent in their method of selection should be noted. The group which exceeded expectations naturally tends to contain those who had rather low L-scores and yet managed to do fairly well on the Latin test. On the other hand, the group which fell short is weighted with some students with extremely high L-scores who still performed very well, though not so outstandingly, on the Latin test.

In an attempt to gain some general picture of the two groups, the Rorschach scores were averaged out into a single profile for each. This procedure is, of course, of doubtful validity since such merging of scores destroys the important interrelations between elements in each individual configuration. Nonetheless, it was adopted as a preliminary expedient to see whether two different patterns would emerge. Fairly clear differences did appear. But it was plain from further examination, particularly of the usually successful group, that analysis by subgroups within each one would produce a more exact and meaningful analysis.

---

with a mean of 20 and a sigma of 4. All scores reported here for both psychological and Latin tests are on this base.

Within the group showing gain, three subgroups can be discerned. The first is made up of the one student of the rather rigid, meticulous type, who was in the first group of 'compulsives' and three other students of this same general type—these three were, in fact, among those cases originally considered for inclusion in the compulsive group but were passed over in favor of slightly more extreme cases. Their presence among the students who actually show a performance better than prediction gives slight substantiation to the original hypothesis which recognized that 'a touch' of rigidity and meticulousness might have a good effect on foreign language learning, but that too great constriction, brought about by too much control, might result in ineffectiveness. Thus these four students, one-third of all the gainers, have patterns marked by the elements mentioned in footnote [3] and give some support to the original hypothesis.

The second subgroup among the students who better expectations suggests a rather different explanation. They are well-adjusted, fairly mature for their age, and working well up to capacity—with adequate resources to handle such adolescent difficulties as they have. It seems scarcely surprising that such students, especially with personal interest and adequate training in a subject, might work at an unusually high level of performance as compared with students generally.

The third and last subgroup among those students who surpass prediction also consists of four students. It seems if they may best be categorized as cases of sublimation. One boy may be taken as the extreme example. Though he exceeds expectation by the minimum margin, his Rorschach pattern among all forty-six students, is that most indicative of extreme maladjustment, suggesting great inner struggle which has thus far been combated by intellectualization. The three much milder cases suggest the same solution. Considerable inner pressure may be channelized into school work and thus the student forges ahead, at least for the time being, under forced draft.

At the other extreme the 'average pattern' for the group whose performance fell short of expectation is fairly representative of all the configurations which form it and, consequently, we need not trouble with subgroups. In general, all these patterns are marked by considerable emotionality, inner conflict, and anxiety.[6]

[6] In the Rorschach scores these qualities appear in the higher scores for C, sum C, m, and K.

They also show a greater tendency to spontaneity and phantasy but, in some cases, these impulses are apparently brought under rather conscious and rigid control.[7] For three students in the group, however, who had unusually high L-scores and failed to perform so outstandingly on the Latin test, (though they did well) we must remember that their specific training, the nature of the Latin test, and the particular distribution of scores on it may account for their presence in this group. But it is also possible that they are subject to the same hypothesis which seems best to account for the other seven cases. This appears to be that emotional disturbances and inner conflicts may be preventing these students from functioning with maximum effectiveness.

For purposes of prediction, these results do not give the simple, conclusive results which were inherent in the original hypothesis tested. Nonetheless, they are somewhat more suggestive than may appear at first glance. Some of the compulsives make considerable gains.[8] Though not all of them are so successful, it is worth noting that no compulsive falls below the level which was to be expected on the basis of his L-score. The same is true of students with Rorschach patterns indicating good adjustment. Some of these students make great gains, while others, seemingly as well adjusted, do not. Probably factors like personal interest in the subject and similar influences make the difference. Yet once again, no student of this type appears in the group of those whose language performance falls short of their verbal ability. Thus (so far as we can judge from these data) both the compulsive students and the well-adjusted ones are good language risks who will at least perform at the level indicated by their L-scores, though we cannot predict on the basis of the Rorschach pattern those who will surpass this norm.

Although the students with inner conflicts and emotional strains exhibit a contradictory performance—some of them exceeding their L-scores and some falling equally far short-the picture is not utterly confused. A possible distinguishing

[7] They have higher M's and FM's than the other students involved in this study as well as the higher color responses just noted. But in the cases with greater control, the F%-scores mount to the level of the compulsive type already discussed, and even beyond it.

[8] The available data do not offer a clear hypothesis as to what features of the Rorschach configuration may be important in distinguishing between those of this type who are and who are not unusually successful in language work.

characteristic in the Rorschach configurations of those who are unusually successful may be their higher Fc and C′ responses in contrast to those students with patterns which are otherwise similar. The present data suggest that these Fc and C′ scores make the difference, but this hypothesis must be tentative until further evidence is available. If, however, both this hypothesis and the theory that a sublimating ability to direct these internal forces into school work is the factor which enables those slightly disturbed[9] students to accomplish more than would be expected are correct, then relatively high Fc and C′ scores may be major indices of this ability to sublimate and, hence, of linguistic performance which equals or surpasses the L-score.

[9] At this point, particularly, the warning should be given that "disturbed" and similar terms are used only on a relative basis as applied to the Rorschach patterns of this group; hence in most cases "disturbed" is more closely synonymous with "less well adjusted" than it is with any expression implying considerable difficulty.

BOOK REVIEWS

CONSTANCE M. MCCULLOUGH, RUTH M. STRANG AND ARTHUR E. TRAXLER. *Problems in the Improvement of Reading.* New York: McGraw-Hill Book Co., 1946, pp.406.

Undoubtedly, teachers and administrators need help with the reading problems they encounter in their schools. The authors consider that this can best be done by a book which includes detailed descriptions of procedure, a statement of principles derived from experiment as basis for the procedures, and examples of approaches to common reading problems. So they have attempted to write a book combining the administrative, guidance and reading technique approaches. The book is organized on a problem basis. Although occasional mention is made of reading in the elementary school and college, the discussion deals primarily with junior and senior high school reading.

Several introductory chapters deal with such topics as present practice, nature of reading, reading process in different situations, growth in reading, and faults that produce failure in reading. The remaining chapters are concerned with problems such as improvement of reading in the school as a whole, reading in the subject fields, and special reading difficulties. Several appendices give information on reading materials and other aids to teaching of reading.

There are numerous contributions in this book that might be cited with warm approval. Among these are the following: (1) Of the many good chapters, the discussions of "Improvement of Reading in the School as a Whole" and "Specific Reading Difficulties" are outstanding. (2) The teaching of reading is the concern of all teachers at all grade levels. (3) Rate of comprehension is a more satisfactory concept than rate of reading. There is no one best rate of comprehension. For the effective reader the rate will vary with the reading purpose and with variations in the comprehending functions. (4) Neither oral nor silent reading should be neglected. Each has a proper place in the reading program. (5) Limitations of standardized tests are discussed. The teacher-made test of reading has an important rôle in any reading program. (6) The citation of case studies yields much pertinent information. (7) Stress is placed

upon guidance in teaching reading. The student is aided in discovering his needs and in developing an attitude receptive to the indicated corrective training. (8) Individualized instruction is emphasized. It is best to take the student where you find him and guide him as far as he can go in improving his reading. (9) Too frequently the kinesthetic method has been ignored while an undue stress has been placed upon either the visual or the auditory approach in teaching reading. A well-balanced program excludes none of these.

Only a few criticisms of this book suggests themselves: (1) Greater emphasis might well have been placed upon reading as a thinking process. (2) Any adequate developmental reading program must give a prominent rôle to reading readiness. Although reading readiness is mentioned, its important relationship to developmental reading is not discussed. (3) The authors' views on interpretation of eye-movement records are sound. They fail to recognize, however, that the study of eye movements is a research technique which has no important use in the reading clinic. (4) The authors place considerable emphasis upon discerning small words in larger ones as a word recognition technique. This technique is of doubtful value except where compound words are involved. In fact, attempts to find little words in big ones frequently lead to errors rather than success. (5) Too little space is devoted to improvement of reading interests and too few citations of reading materials are given. (6) The effectiveness of the book would have been increased if only high school and college reading problems were considered. Certain material is introduced which is relevant only to problems in the elementary school.

This text is an important contribution to the field. It is a well-written, highly practical treatise. Every teacher and administrator in junior and senior high schools should be familiar with its contents.

MILES A. TINKER
University of Minnesota

S. R. SLAVSON. *Recreation and the Total Personality.* New York: The Association Press, 1946, pp. 205.

It seems extremely likely, almost inevitable, that men and women in most parts of the world will in the future be able to

support themselves by working a decreasing number of hours each week. A large fraction of the waking time not devoted to remunerative work is and will be spent in recreation. This book indicates how recreational activities can make a maximum contribution to the personality adjustment and general well-being of men and women as well as, boys and girls.

In his initial chapter Slavson makes distinctions between recreation, drudgery, labor, work, and play. He recognizes that his definitions are somewhat empirical and contends that "recreation does not consist of what one does; it is rather the motive, attitude, and value of the doing to the individual that gives the activity a recreational significance." (p. 2) He conceives of a subtle but important difference between play and recreation. The latter serves more completely the need for relaxation from effort and tension, and can bring into psychological and physical balance those mental and physical activities that are apt to be neglected in the routine of daily activities. Recreation is thought of as being 're-creation'.

Slavson believes that recreation serves different purposes for different people, and is, by its very nature, pleasurable because it satisfies some craving or need. This satisfaction may range from exercising a set of stagnant muscles to the need for building up one's ego by overcoming a contestant in some game. While there is much that we do not know about motivation, Slavson's explanation of the relationship between the recreational activities we chose and our 'unconscious motivations' was clear and persuasive. Very little objective data were advanced in support of this argument, but it's plausibility was impressive. Recreation, in large part at least, provides a method for balancing the psychological deprivation from which we all suffer—deprivation such as physical underfunctioning, the constant frustration of unconscious drives, devaluated self-esteem, lack of self-assertation, and inadequate creative expression.

In connection with his discussion of the relationship of recreation to mental hygiene, Slavson contends that "man's psychological drives and needs are immutable. They cannot be destroyed or repressed, but their expression can be transformed and changed through definite, knowable processes." (p. 27) He contends furthermore that present social institutions rather con-

sistently oppose the natural expression of many of our needs. One of the major advantages of recreation is that the individual can express some of these needs without feeling guilt because the recreational method of expression is by and large acceptable to society.

Slavson's discussion of the function of recreation is developed within the frame of reference of a democratic society. He believes that the recreational pattern in an autocratic state would be different from that which should be encouraged in a democratic society. He insists that "recreation as individual release and as groups experience can flourish only in a democracy." (p. 74) The reviewer dissents, although the variety of 'free choice' recreation in a non-democratic state is usually reduced.

While the part of the book that interested the reviewer most had to do with the relationship between the dynamics of psychology and recreation, Slavson also deals with many topics of more immediately practical significance. He discusses competition, the rôle of the leader in recreation, the rôle of the supervisor, and the community resources that can, if available, help make recreation therapeutically more valuable.

Slavson's book is more than a treatise on recreation. It is in large part a discussion of the dynamics of human personality development. As such, the book can be read with benefit by anyone who is interested in ways and means of releasing some of the tensions that are characteristic of our culture, and damaging to normal personality development.

STEPHEN M. COREY
The University of Chicago.

WERNER WOLFF. *The Personality of the Preschool Child.* New York: Grune and Stratton, 1946, pp. 341.

Studies of child behavior have of necessity been made by adults; and this fact alone has introduced a frequently unrecognized variable in the total experimental situation. Of perhaps greater importance is the influence of adult reasoning in the interpretation of observed child behavior. Thus children's thought has been considered primitive and illogical, their drawings mean-

ingless or inchoate, their play happy and without care, and their behavior growth a matter of increasingly complex reflex conditioning. During the past decade or two there has been an increasing recognition that the child must be studied in his total behavior, and the interpretation of finding must be made in terms of child, and not adult, standards. Dr. Wolff has been one of the foremost proponents of this view. His theoretical position as developed in this book depends upon "the recognition of two worlds in which young child and adult live isolated from each other," and on the concept that the behavior expressions of the young child are variants of the "child's search for his self."

The present study has three main divisions. In the first the author discusses the child's thinking, emotions, social life, and concepts of reality as they compare to those of the adult, and as they may be interpreted from observations of behavior. The second part, which constitutes over one half of the book, is devoted to a discussion of the author's experiments with preschool children, particularly the interpretation of their drawing. Here, in abundant illustration, the author shows that even the crudest man, or the simplest scribbling, is not without meaning when judged as expressive behavior of the child. Theoretical discussions of children's art, of the 'educational bridge' between the worlds of the child and the adult, and of methods of investigation in child psychology, constitute the third major division. It should also be noted that Wolff is not dependent upon his own investigations alone; a bibliography of six hundred twenty-eight titles, most of which are pertinently referred to in the text, is adequate evidence of comprehensive scholarship.

The possibilities in the author's method, called 'experimental depth psychology,' appear to be great. The methods and theory are not psychoanalytically oriented, nor are they limited to behavioristic atomism. Wolff earnestly tries to study the behaving child, and to interpret his findings with parcimony. Critics may point out that much of the work depends upon a single form of behavior—drawing. However, the author introduces other evidence with sufficient frequency to indicate that his interpretations would follow equally well from other expressive behavior. Scientists interested in the study of child behavior cannot afford

to miss this book, and teachers and parents will profit a great deal, especially from its point of view. C. M. LOUTTIT
Sampson College

HELEN M. ROBINSON. *Why Pupils Fail in Reading.* Chicago: The University of Chicago Press, 1946, pp. 257.

This book depicts an ambitious attempt to: (1) summarize, interpret and evaluate the findings and opinions of reading authorities concerning causes of severe reading retardation; (2) study with the coöperation of a group of specialists thirty poor readers; (3) present conclusions and new problems which the study has found and pointed out.

The techniques used in the study were somewhat unique in that not only were specialists utilized in making the examinations of each of the thirty children studied, but their coöperation was also enlisted in a conference "for the purpose of looking at the child as a whole and evaluating each separate finding in the light of the total picture." Following the carrying out of the group's first recommendations, a series of group conferences was held to review cases and make additional recommendations in the light of developments in various cases after the last group conference.

The specialists including a social worker, a psychiatrist, a pediatrician, a neurologist, three ophthalmologists, a speech-correction specialist, an otolaryrgologist, an endocrinologist, a reading specialist, and the investigator, who acted as psychologist and reading technician, who worked together on the thirty cases over a period of five years.

Main conclusions as summarized by Dr. William S. Gray in the Preface are: (1) "pupils who are seriously retarded exhibit many anomalies, that is, physical, mental, social, and emotional deficiencies or disturbances"; (2) "as a rule, the number of anomalies is greater, the more serious the retardation in reading"; and (3) "many of the anomalies exhibited have little or no relation to reading retardation."

Study of basic motivation and methods of teaching as possible causes of poor reading are given little attention. This limitation

of the study is recognized by the author and the following explanation is given: "At the time of the examination it was not possible to evaluate the methods of teaching reading which had been used on each child in the years before examination." Since motivation and method are probably the areas in which the teacher is in a position to make the greatest changes in trying to eliminate the causes of poor reading, the limitation of the study in this regard is serious from the standpoint of the classroom teacher. The treatment aspects of the study suggests some procedures for the teacher to use after the damage has been done. Some of the dynamics of the cases are partially covered in the findings of the psychiatrist and social worker.

It is the opinion of the reviewer that this book is a very valuable one for the reading specialist and the person doing research in reading improvement. The summary of previous experimental results, the picture of the techniques used in this study, and the results of the study are likely to prove of interest and value to the professional worker who has had considerable training in the field.

The book is designed to analyze possible causal factors and is not intended to be one especially for the classroom teacher who wants specific suggestions 'he will feel it is feasible to use' in improving the reading abilities of his learners. Although Mrs. Robinson says that "the results of this study are very suggestive to anyone interested in poor readers," it is very doubtful whether this would be the case with many who are 'interested in poor readers.' In fact, to many parents and to many teachers, and to most poor readers themselves, all of whom are 'interested in poor readers', the results would probably prove quite confusing. Many professional persons with advanced training who are interested in poor readers will find the book of interest.

RAY H. SIMPSON

University of Alabama

KENNETH B. HAAS AND HARRY Q. PACKER. *Preparation and Use of Visual Aids.* New York: Prentice-Hall, 1946, pp. 224. $4.00.

It is suggested that this is a suitable text for courses in the technique and methods used in the preparation and use of visual

aids. In each chapter is a practical presentation of a visual aid. Details are given on how to use, how to prepare, and where to find these aids. Among the aids included are motion pictures, stripfilm, slides, opaque projector, maps and diagrams, flash cards, posters and manuals, pictures, blackboard and others. The text is intended to meet the needs of those who can use visual aids in education and in many other situations.

The discussions are simple and clear. The lists of sources of materials will be found useful and complete. The educator who employs visual aids in his teaching will find this book to be decidedly useful. Nevertheless, the authors' suggestion that twenty-five to thirty-five per cent more factual information will be gained and thirty-five per cent more of that information retained if visual aids are employed is perhaps over-optimistic.

MILES A. TINKER

University of Minnesota

MARY FORD. *The Application of the Rorschach Test to Young Children.* Minneapolis: University of Minnesota Press, 1946, pp. 114.

In this monograph, #23 of the University of Minnesota Institute of Child Welfare, Dr. Ford reports her investigations on the responses of children from three to eight years of age to the Rorschach test. The reader cannot help but be impressed at the evident objectivity and care with which this research has been planned and carried out. There were four major aims of the investigation: (1) the refinement and simplification of procedures adapting them to young children, (2) the study of variations in determinants as related to age, intelligence and sex, (3) the determination of reliability of determinants at young age levels, and (4) a check on the validity of claims as to the meaning of certain test determinants.

The one hundred twenty-three children examined were attending nursery or elementary schools in Minneapolis and Ithaca, New York. Except for a restriction on rotating the cards, the administrative procedures were standard. There was no evi-

dence that this one restriction influenced the responses. The results indicate that the Rorschach is definitely useful at the young ages studied; the children utilized all the ways of responding which are utilized by adults, but not to the same extent. The percentage rejecting at least one card decreased rapidly from thirty-two per cent at three years of age to none at six years of age. The monograph reviews the literature, especially of children, involved with work with inkblots. The procedures and results of the investigation are presented in detail, and illustrated with both tables and graphs. It is hardly necessary in this review to attempt to summarize the details. The important conclusions are that the test proves useful with children, although there is a great need for norms at young ages. The reliabilities varied, with some determinants showing high correlations between test and retest, and with others being low. The results with these young children do not always support certain validity claims based on work with adults.

This study is an important contribution to the Rorschach literature. A particularly significant contribution for the Rorschach layman is a glossary of terms and symbols used in Rorschach interpretation. There is a bibliography of one hundred forty-four titles.

C. M. Louttit

Sampson College

Edgar Dale. *Audio-visual Methods in Teaching.* New York: The Dryden Press, 1946, pp. 546. $4.25.

The emphasis upon the use of auditory and visual materials in teaching has increased greatly during recent years. At the same time the availability of these materials has shown a marked improvement. The present volume is more comprehensive than any other treatise in the field. It was planned both as a textbook and as a tool for teachers in service. Its three major sections deal with: (1) theory of audio-visual materials, (2) audio-visual teaching materials, and (3) audio-visual methods applied in the classroom and in the school system. Useful lists of source materials are given.

The materials in this text are comprehensive, clearly presented and interesting. Every teacher who wishes to take advantage of auditory and visual aids in his work cannot afford to be without a copy. The book deserves to be widely read and education will be the better for it.

MILES A. TINKER
University of Minnesota